



TESIS-KI142502

Deteksi Anomali menggunakan *Control Flow Patterns* dan *Fuzzy Regression* di Terminal Petikemas

DEWI RAHMAWATI
NRP. 05111650010062

DOSEN PEMBIMBING
Prof. Drs.Ec. Ir. Riyanarto Sarno, M.Sc., Ph.D
NIP. 19590803 198601 1 001

PROGRAM MAGISTER
BIDANG KEAHLIAN MANAJEMEN INFORMASI
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN KOMUNIKASI
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2017

[Halaman ini sengaja dikosongkan]

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Komputer (M. Kom)
di
Institut Teknologi Sepuluh Nopember

Oleh :
Dewi Rahmawati
NRP. 05111650010062

Dengan judul :
DETEKSI ANOMALI MENGGUNAKAN CONTROL FLOW PATTERNS DAN FUZZY
REGRESSION DI TERMINAL PETIKEMAS

Tanggal Ujian : 08-01-2018
Periode Wisuda : 2017 Gasal


Disetujui oleh :

Prof. Drs.Ec. Ir. Riyanarto Sarno, M.Sc., Ph.D.
NIP. 19590803 198601 1 001




(Pembimbing 1)

Dr. Eng. Darlis Heru Murti, S.Kom, M.Kom.
NIP. 19771217 200312 1 001



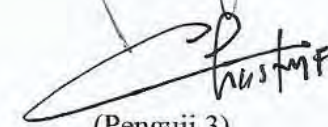
(Penguji 1)

Dr. Ir. Raden Venantius Hari Ginardi, M.Sc.
NIP. 19650518 199203 1 003



(Penguji 2)

Dr. Eng. Chastine Fatichah, S.Kom, M.Kom.
NIP. 19751220 200112 2 002



(Penguji 3)



Dr. Agus Zainal Arifin, S. Kom., M. Kom.
NIP. 19720809 199512 1 001

an. Direktur Program Pascasarjana ITS,
Asisten Direktur Program Pascasarjana ITS,

[Halaman ini sengaja dikosongkan]

Deteksi Anomali menggunakan *Control Flow Patterns* dan *Fuzzy Regression* di Terminal Peti Kemas.

Nama mahasiswa : Dewi Rahmawati
NRP : 05111650010062
Pembimbing I : Prof. Drs.Ec. Ir. Riyanarto Sarno, M.Sc. Ph.D.

ABSTRAK

Kasus kriminal pada terminal peti kemas di Indonesia dapat disebabkan oleh banyak faktor salah satunya adalah anomali dalam model proses. Menanggapi permasalahan ini, penelitian ini mengusulkan metode pendeteksian anomali pada penanganan kontainer pelabuhan untuk mengurangi kerugian dan kewaspadaan terhadap kasus yang terdeteksi anomali. Paramater atau atribut yang digunakan untuk menganalisis anomali adalah Analisis Salah Urutan (*Wrong Pattern Analysis*) dengan metode baru yakni menggunakan metode *Control Flow Pattern* yang akan dianalisis dengan menggunakan *Isomorphism Graph* dan Analisis Urutan Proses Lurus (*Skip Sequence*), Analisis Waktu Operasi Maksimal (*Wrong Throughput Time Max*), Analisis Waktu Operasi Minimal (*Wrong Throughput Time Min*) dan Analisis Salah Keputusan (*Wrong Decision*). Sehingga total atribut yang dijadikan faktor penyebab terjadinya anomali dalam tesis ini sejumlah 5 atribut. Dalam penelitian ini, *event log* diperoleh dari database PT.Terminal Peti Kemas Surabaya kemudian penyeleksian *event log* yang dibutuhkan, setelah proses seleksi terjadi, maka *event log* yang tersaring berdasarkan 3 faktor adalah *noise log*, *incompleteness log* dan *truncated logs*. Hasil log yang difilter digunakan untuk mendeteksi anomali. Kemudian membentuk *Control Flow Patterns* berupa *Anomaly Patterns* yang digunakan untuk membentuk suatu proses model yang dapat digunakan untuk mendeteksi pola yang salah dengan menggunakan *Isomorphisms Graph*. Pendeteksian anomali dilakukan secara otomatis dengan menggunakan program dan hasil deteksi anomali akan dianalisis dengan menggunakan 2 metode yaitu metode *Fuzzy Regression* dan *Multiple Linear Regression* dimana keduanya akan dibandingkan yang mana algoritma terbaik untuk mendeteksi anomali. *Multiple Linear Regression* adalah algoritma untuk mengetahui tingkat prediksi

anomali dari semua nilai atribut anomali sehingga nilainya *continuous* (0,1,0,2, ..., X_n). *Fuzzy Regression* adalah teknik baru untuk mengetahui tingkat prediksi kecurangan yang lebih akurat daripada regresi linier berganda, metodenya adalah mengklasifikasikan kasus mana yang memiliki anomali rendah, anomali sedang dan anomali tinggi kemudian menemukan tingkat prediksi anomali dari semua nilai atribut anomali yang nilainya *discrete* (Rendah, Sedang dan Tinggi). Hasil penelitian menunjukkan bahwa *Fuzzy Regression* lebih tepat untuk mendeteksi anomali daripada *Multiple Linear Regression*. Terbukti dengan sensitivitas dan spesifisitas *Fuzzy Regression* untuk anomali yang tinggi adalah 78% dan 99%, sensitivitas dan spesifisitas *Fuzzy Regression* untuk anomali medium adalah 85% dan 99%, sensitivitas dan spesifisitas *Fuzzy Regression* untuk anomali rendah adalah 88% dan 99% dan sensitivitas dan spesifisitas *Fuzzy Regression* untuk tidak anomali adalah 99% dan 86%.

Kata kunci: Anomali, *Fraud*, *Anomaly Patterns*, *Event Logs*, *Control Flow Patterns*, *Fuzzy Regression*, *Multiple Linear Regression*.

Anomaly Detection using Control Flow Pattern and Fuzzy Regression in Port Container Terminal.

Name : Dewi Rahmawati
Student Identity Number : 05111650010062
Supervisor : Prof. Drs.Ec. Ir. Riyanarto Sarno, M.Sc. Ph.D.

ABSTRACT

A criminal case on port container handling in Indonesia can be caused by many factors. One of them is an anomaly in the process model. Responding to the issue this research proposes a method to detection anomaly on port container handling for decrease the loss of port container handling and to be wary on a case that detected an anomaly. In order to be able to detect anomalies on the PT. Terminal Container Surabaya is based on the 5 Attributes are Skip Sequence, Wrong Throughput Time Max, Wrong Throughput Time Min, Wrong Decision and Wrong Pattern Analysis. In this research, event logs are obtained from the database PT. Terminal Container Surabaya through the process of selecting the required event log, after the selection process occurs, then the process of filtering event log based on 3 factors are noise logs, incompleteness logs, and truncated logs. The result of the filtered log is used to detect anomalies. Then build a Control Flow Patterns for form a model process that can be used for a detected the wrong pattern using isomorphism graphs called anomaly patterns. Anomaly detection is done automatically by using the program and the results of the anomaly detection will be analyzed by using two methods are Fuzzy Regression method and Multiple Linear Regression where the two will be compared which is the best algorithm to detect an anomaly. Multiple linear regression is an algorithm to find predicts the rate of the anomaly from the all of anomaly attribute value that the value is continuous (0.1, 0.2,... Xn). Fuzzy regression is the new technique to find the predicts rate of fraud that more accurately than multiple linear regression, the step are first classified which case having no anomaly, low anomaly, medium anomaly and high anomaly then find the predicts rate of anomaly from the all of anomaly attribute value that the value is discrete (No Anomaly, Low Anomaly, Medium Anomaly and High

Anomaly). The result shows that fuzzy regression is more appropriate for detecting anomaly than multiple linear regression. Proved by the result of sensitivity and specificity are 78% and 99% for the high anomaly, 85% and 99% for the medium anomaly, 88% and 99% for low anomaly and 99% and 86% for no anomaly.

Keywords: *Anomaly, Fraud, Anomaly Patterns, Event Logs, Control Flow Patterns, Fuzzy Regression, Multiple Linear Regression.*

KATA PENGANTAR

Penulis mengucapkan rasa syukur yang tak berhingga kepada Allah SWT atas segala rahmat, berkah, hidayah, kesehatan dan petunjuk-Nya, sehingga penulis dapat menyelesaikan tesis yang merupakan salah satu syarat dalam menyelesaikan Program Studi Magister di Institut Teknologi Sepuluh Nopember Surabaya.

Terselesaikannya tesis beserta laporannya ini tentunya tak luput dari peran serta berbagai pihak yang telah memberikan bantuan dan dorongan semangat, baik secara langsung maupun tak langsung. Untuk itu, atas segala bantuan yang telah diberikan, penulis mengucapkan terima kasih serta penghargaan yang sebesar-besarnya antara lain kepada:

1. Kedua Orang Tua penulis yang senantiasa memberikan motivasi, semangat, dan harapan serta mendoakan penulis demi keberhasilan penulis dalam menyelesaikan studi.
2. Bapak Prof. Drs.Ec. Ir. Riyanarto Sarno, M.Sc., Ph.D M.Kom. selaku dosen pembimbing yang telah banyak meluangkan waktu, tenaga dan pikiran dalam membimbing penulis sehingga tesis ini dapat terselesaikan dengan baik, serta membuka wawasan penulis akan luasnya ilmu pengetahuan. Terima kasih banyak atas segala kebaikan Bapak, semoga Allah SWT senantiasa merahmati Bapak dan keluarga.
3. Bapak Waskitho Wibisono, S.Kom., M.Eng., Ph.D. selaku ketua Program Studi Program Pascasarjana Teknik Informatika yang telah banyak memberikan bantuan dalam memecahkan masalah akademik.
4. Ibu Dr. Eng. Chastine Fatichah, S.Kom, M.Kom., Dr. Eng. Darlis Heru Murti, S.Kom, M.Kom, Dr. Ir. Raden Venantius Hari Ginardi, M.Sc selaku dosen penguji yang telah banyak membantu penulis untuk bisa menjadi lebih baik.
5. Seluruh dosen S2 Teknik Informatika yang telah memberikan ilmu dan pengetahuan kepada penulis selama menempuh studi.
6. Semua rekan mahasiswa S2 Teknik Informatika utamanya angkatan 2016 yang bernama Kelly Rossa Sungkono, Abdul Charis Fauzan, Afina Lina Nurlaili dan Yutika

Amelia Effendi selaku rekan seperjuangan yang telah memberikan bantuannya baik secara langsung maupun tidak langsung.

Semoga Allah SWT memberikan balasan yang setimpal dan selalu memberikan taufik serta hidayah-Nya bagi semua pihak yang telah banyak memberikan bantuan, bimbingan ataupun nasehat-nasehat.

Penulis menyadari bahwa dalam laporan tesis ini masih banyak kekurangan. Karena itu, masukan ataupun saran demi perbaikan dan penerapan tesis ini dimasa mendatang tetap penulis harapkan.

Surabaya, 8 Januari 2018

Penulis

DAFTAR ISI

| | |
|--|------|
| ABSTRAK | v |
| ABSTRACT | vii |
| KATA PENGANTAR | ix |
| DAFTAR ISI | xi |
| DAFTAR GAMBAR | xiii |
| DAFTAR TABEL | xv |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Perumusan Masalah | 4 |
| 1.3 Tujuan Penelitian | 5 |
| 1.4 Batasan Masalah | 5 |
| 1.5 Kontribusi Penelitian | 5 |
| 1.6 Manfaat Penelitian | 5 |
| 1.7 Sistematika Penulisan | 5 |
| BAB II DASAR TEORI DAN KAJIAN PUSTAKA | 7 |
| 2.1 Deteksi Anomali dan Fraud | 7 |
| 2.2 <i>Process Mining for Process-Based Fraud Detection</i> | 12 |
| 2.3 <i>Atribut Process Based Fraud (PBF)</i> | 13 |
| 2.4 <i>Data Mining</i> untuk Deteksi Fraud | 16 |
| 2.5 <i>Association Rule Learning Applied</i> untuk Deteksi Fraud | 17 |
| 2.6 <i>Graph Pattern Matching</i> | 19 |
| 2.7 Graf Isomorfisma | 21 |
| 2.8 <i>Control Flow Patterns</i> | 23 |
| 2.9 <i>Multiple Linear Regression</i> | 27 |
| 2.10 <i>Fuzzy Regression</i> | 29 |
| BAB III METODE PENELITIAN | 38 |
| 3.1 Sumber Data Penelitian | 42 |

| | | |
|--|---|----|
| 3.2 | Pembentukan <i>Control-Flow Pattern</i> dengan rule <i>anomaly patterns</i> | 42 |
| 3.3 | Pendeteksian Wrong Patterns dengan Rule Anomaly Patterns | 43 |
| 3.4 | Pendeteksian Anomali dengan Program Java | 45 |
| 3.5 | Pendefinisian <i>Attribute Values</i> | 48 |
| 3.6 | Data Penelitian untuk Proses <i>Fuzzy Regression</i> | 50 |
| BAB IV UJI COBA DAN ANALISIS HASIL | | 57 |
| 4.1 | Evaluasi Model dengan <i>Multiple Linear Regression Model</i> | 57 |
| 4.2 | Evaluasi Model dengan <i>Fuzzy Regression Model</i> | 66 |
| BAB V KESIMPULAN | | 77 |
| DAFTAR PUSTAKA..... | | 79 |
| DAFTAR SINGKATAN..... | | 83 |
| DAFTAR ISTILAH..... | | 85 |
| DAFTAR INDEKS..... | | 91 |
| BIOGRAFI PENULIS | | 93 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2.1. Pattern and Data Graph | 21 |
| Gambar 2.2. Contoh graf isomorfis | 22 |
| Gambar 2.3. Contoh graf tidak isomorfis | 22 |
| Gambar 2.4. Sequence Patterns | 25 |
| Gambar 2.5. <i>Parallel Split Patterns</i> | 25 |
| Gambar 2.6. <i>Synchronization pattern</i> | 26 |
| Gambar 2.7. <i>Exclusive choice pattern</i> | 26 |
| Gambar 2.8. <i>Simple merge pattern</i> | 27 |
| Gambar 2.9. Model Residual Standarisasi untuk for X_i - X_n | 28 |
| Gambar 2.10. Koefisien Fuzzy | 30 |
| Gambar 2.11 Konseptualisasi upper dan lower bound..... | 32 |
| Gambar 2.12. Interval Fuzzy Regresi..... | 32 |
| Gambar 2.13. Parameter fuzzy simetris | 33 |
| Gambar 2.14. Estimasi A_j menggunakan faktor “ <i>h-certain</i> ” | 34 |
| Gambar 2.15. Pengamatan Keluaran Fuzzy..... | 35 |
| Gambar 2.16. Keluaran Fuzzy dari Model..... | 36 |
| Gambar 2.17. Fitting perkiraan output ke output yang diamati | 36 |
| Gambar 2.18. Model Fuzzy Linear Regression dan model <i>h-certain</i> | 37 |
| Gambar 3.1 Desain Sistem..... | 39 |
| Gambar 3.2. Desain Sistem (Lanjutan)..... | 40 |
| Gambar 3.3. Bagan gambaran proses input output pembentukan anomaly patterns | 42 |
| Gambar 3.4. <i>Build Control Flow Patterns</i> | 43 |
| Gambar 3.5. <i>Rules Anomaly Pattern</i> | 43 |
| Gambar 3.6. Alur Pendeteksian <i>Wrong Patterns</i> dengan <i>Rule Anomaly Patterns</i> | 44 |
| Gambar 3.7. Contoh Penerapan <i>Isomorphisms Graph</i> pada Case | 44 |
| Gambar 3.8. Contoh Penerapan <i>Isomorphisms Graph with dependencies matrix</i> pada Case mengandung anomali | 45 |

| | |
|---|----|
| Gambar 3.9. Contoh Penerapan <i>Isomorphisms Graph with dependencies matrix</i> pada Case tidak mengandung anomali..... | 45 |
| Gambar 3.10. <i>Pseudocode for Skip Sequences and Skip Decision</i> | 46 |
| Gambar 3.11. <i>Pseudocode for Throughput Time Min/Max</i> | 46 |
| Gambar 3.12. <i>Pseudocode for the Wrong Pattern</i> | 46 |
| Gambar 3.13. <i>Pseudocode for Wrong Decision</i> | 46 |
| Gambar 3.14. <i>Program Detection Wrong Patterns</i> | 47 |
| Gambar 3.15. <i>Program Detection Input Data Time for Detection Wrong Throughput Time Min/Max</i> | 47 |
| Gambar 3.16. <i>Program Detection Wrong Throughput Time Min/Max</i> | 48 |
| Gambar 3.17. <i>Program Detection Wrong Decisions</i> | 48 |
| Gambar 3.18. Interval Fuzzy Regression | 51 |
| Gambar 3.19. Gambaran <i>Business Process</i> Terminal Peti Kemas (TPS) | 52 |
| Gambar 3.20. Pembentukan Model Proses dengan Yawl Petri Net | 53 |
| Gambar 4.1. <i>MLR Model of ROA (Rate of Anomaly) High Anomaly</i> | 64 |
| Gambar 4.2. <i>MLR Model of ROA (Rate of Anomaly) Medium Anomaly</i> | 64 |
| Gambar 4.3. <i>MLR Model of ROA (Rate of Anomaly) Low Anomaly</i> | 64 |
| Gambar 4.4. <i>Plot Membership Function Fuzzy Rate of Anomaly (ROA) High Anomaly</i> .. | 73 |
| Gambar 4.5. <i>Probability Plot of Detecting High Anomaly in Fuzzy High Anomaly</i> | 73 |
| Gambar 4.6. <i>Sensitivity Value of FR and MLR</i> | 76 |
| Gambar 4.7. <i>Specificity Value of FR and MLR</i> | 76 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1. Penelitian Deteksi Fraud Untuk Proses Bisnis | 9 |
| Tabel 2.2. Sepuluh Atribut untuk Mendeteksi Fraud (Huda, 2015) | 14 |
| Tabel 2.3. Penjelasan Atribut Deteksi Fraud | 15 |
| Tabel 2.4. Metode <i>Data Mining</i> untuk Deteksi Fraud | 19 |
| Tabel 2.5. <i>Control Flow Pattern</i> | 24 |
| Tabel 2.6. <i>Data Pairs</i> | 31 |
| Tabel 3.1. <i>Attribute Value Skip Sequences (31 Activity)</i> | 49 |
| Tabel 3.2. <i>Attribute Value Throughput Time Min</i> | 49 |
| Tabel 3.3. <i>Attribute Value Throughput Time Max</i> | 49 |
| Tabel 3.4. <i>Attribute Value Wrong Decisions</i> | 49 |
| Tabel 3.5. <i>Attribute Value Wrong Patterns</i> | 50 |
| Tabel 3.6. Sumber Database 65000 Log (41 kolom label) PT.Terminal Peti Kemas Surabaya..... | 54 |
| Tabel 3.7. Sumber Database 600 Log (12 kolom label) PT.Terminal Peti Kemas Surabaya telah terseleksi siap dideteksi | 55 |
| Tabel 3.8. Tabel Data Penelitian Rate of Anomaly (From Expert) | 56 |
| Tabel 4.1. Tabel <i>Attribute Value</i> dan <i>Rate of Fraud</i> | 57 |
| Tabel 4.2. <i>Category of Anomaly in Multiple Linear Regression</i> | 58 |
| Tabel 4.3. <i>Number of Cases of Training and Testing Multiple Linear Regression</i> | 58 |
| Tabel 4.4. <i>Data Training ROAHIGH</i> | 58 |
| Tabel 4.5. <i>Data Training ROAMED</i> | 59 |
| Tabel 4.6. <i>Data Training ROALOW</i> | 60 |
| Tabel 4.7. <i>Data Training ROANo</i> | 60 |
| Tabel 4.8. <i>Testing Data High Anomaly</i> | 62 |
| Tabel 4.9. <i>Confusion Matrix for Multiple Linear Regression</i> | 65 |
| Tabel 4.10. <i>Table Fuzzy Skip Sequences</i> | 66 |
| Tabel 4.11. <i>Table Fuzzy Wrong Throughput Time Min</i> | 67 |

| | |
|--|----|
| Tabel 4.12. <i>Table Fuzzy Wrong Throughput Time Max</i> | 67 |
| Tabel 4.13. <i>Table Fuzzy Wrong Decisions</i> | 67 |
| Tabel 4.14. <i>Table Fuzzy Wrong Pattern</i> | 67 |
| Tabel 4.15. <i>Table Fuzzy Rate of Anomaly</i> | 67 |
| Tabel 4.16. <i>Low Anomaly of 5 Attributes</i> | 68 |
| Tabel 4.17. <i>Medium Anomaly of 5 Attributes</i> | 68 |
| Tabel 4.18. <i>High Anomaly of 5 Attributes</i> | 68 |
| Tabel 4.19. <i>No Anomaly of 5 Attributes</i> | 68 |
| Tabel 4.20. <i>Testing Data High Anomaly Fuzzy Regression</i> | 70 |
| Tabel 4.21. <i>Penggolongan Kategori Fuzzy</i> | 72 |
| Tabel 4.22. <i>Number of Cases of Training and Testing Fuzzy Regression</i> | 73 |
| Tabel 4.23. <i>Confusion Matrix for Calculate Sensitivity Fuzzy Regression</i> | 74 |
| Tabel 4.24. <i>Confusion Matrix for Calculate Sensitivity Fuzzy Regression</i> | 75 |
| Tabel 4.25. <i>Result Detection Anomaly using Fuzzy Regression</i> | 75 |

BAB I

PENDAHULUAN

1.1 Latar Belakang

Anomaly merupakan suatu penyimpangan atau keanehan yang sering terjadi pada suatu proses bisnis. Anomali juga sering disebut sebagai suatu kejadian yang tidak bisa diperkirakan sehingga sesuatu yang terjadi akan berubah dari kejadian biasanya. Anomali bisa menjadi suatu indikasi adanya kejadian kecurangan dalam perusahaan. Anomali adalah bentuk kejahatan yang mengambil keuntungan dari berbagai modus kecurangan. Anomali telah menjadi perhatian yang signifikan karena merupakan penyebab utama kerugian dalam organisasi dan perusahaan (Amara, Amar and Jarboui, 2013).

Perusahaan berpotensi mengalami kerugian finansial karena proteksi anti-anomali mereka tidak mampu mendeteksi semua kasus anomali. Ada kemungkinan untuk mendeteksi anomali jika sistem peringatan dini berjalan dengan baik. Misalnya, penyimpangan dalam prosedur operasi standar (SOP) yang dilakukan oleh anggota staf dapat dideteksi lebih awal, sehingga perusahaan dapat memodifikasi pola kerja anggota staf untuk mengurangi kemungkinan anomali. Dalam kasus tersebut, *process mining* memberikan solusi dengan memberikan opsi untuk menguji kesesuaian proses bisnis dengan SOP [Kalyani and Devi, 2011].

Data mining dan deteksi fraud telah dipelajari selama beberapa dekade dengan berbagai cara (Ngai, Hu, Wong, Chen and Sun, 2010). Sebagai contoh, dengan menggunakan algoritma jaringan syaraf tiruan algoritma *self organizing maps*, teori *dempster-shafer* dan algoritma pembelajaran bayesian, model klasifikasi, analisis empiris, dan kolaborasi *web service*. Selain itu, dalam *process mining* telah dilakukan analisis pengendalian arus, analisis peran dan analisis kinerja, association rule learning, hibrida ARL dan *process mining*.

Deteksi anomali sangat penting untuk meminimalkan kesalahan yang terjadi pada perusahaan (Shen, Tong and Deng, 2007). Hal Ini harus diidentifikasi dalam proses bisnis yang dapat dianalisis dengan process mining, termasuk *performance*, *event sequence*,

control flow dan *role analysis*. Deteksi dilakukan dengan menggunakan *data mining* (yaitu *association rule learning*) dan kombinasi *data mining* dan *process mining* (metode hibrida), yang hasilnya kemudian dianalisis berdasarkan proses bisnis masing-masing untuk mengidentifikasi penyimpangan SOP.

Process mining telah diusulkan dalam (Chae, Shime, Cho and Lee, 2007) untuk mengurangi anomali dengan menggunakan analisis kinerja, analisis aliran kontrol, dan analisis peran untuk mempelajari proses bisnis. Tetapi metode ini tidak termasuk algoritma untuk deteksi anomali. Telah terbukti bahwa *process mining* mampu mendeteksi anomali dalam proses bisnis. Konsep 1 + 5 + 1 mengusulkan alat untuk implementasi deteksi PBF. singkatan dari "1 + 5 + 1" adalah (1) persiapan log + (5) {1} analisis log, {2} analisis proses, {3} analisis kesesuaian, {4} analisis kinerja, {5} analisis sosial menggunakan *filter*, *summarization*, *sorting*, *join* dan *aging* + (1) iterasi dan *refocusing* (Stoop, 2012).

Algoritma *rules learning* (ARL) telah digunakan untuk menganalisis korelasi antara aturan anomali dengan perilaku transaksi kartu kredit (Huda, Sarno, Ahmad and Santosa, 2014). Perilaku originator (pengguna yang mengeksekusi suatu acara) yang konsisten dengan karakter perilaku anomali curang ditandai sebagai anomali actor. Penelitian ini mendeteksi kecurangan dengan menganalisis penyimpangan SOP dengan menggunakan metode *non-fuzzy*. Oleh karena itu, nilai kondisinya (yaitu *not anomaly*, *between not anomaly and anomaly*, *anomaly*, *confident anomaly* and *very confident anomaly*) tidak ditentukan..

Dalam studi lain (Sarno, Dewandono, Ahmad, Naufal and Sinaga, 2015), metode hibrid menggabungkan algoritma ARL dan *process mining* diusulkan. Algoritma ARL digunakan untuk mengidentifikasi perilaku anomali, sedangkan *process mining* digunakan untuk menganalisis penyimpangan SOP. Metode yang digunakan adalah pendapat ahli tentang *association rule learning* untuk menghasilkan peraturan pemeriksaan kepatuhan. Penelitian ini mempertimbangkan bobot atribut, yang ditentukan secara subyektif, untuk menentukan kecurangan.

Namun, penelitian terdahulu ini hanya mempertimbangkan kondisi *non-fuzzy* dan metode hibrid dalam mendeteksi anomali dan hanya mendeteksi satu aktivitas yang

memiliki atribut. Dalam thesis ini mengusulkan agar anomali dapat dideteksi berdasarkan Graph Pattern Matching yakni menggunakan control flow patterns yang dapat mendeteksi anomali lebih baik dengan mengukur seberapa banyak tingkat anomali dalam suatu case dalam kelompok atau yang disebut konspirasi antara pelaku anomali. Proses bisnis yang digunakan untuk mendeteksi kecurangan adalah proses bisnis *scalable* yang pada kenyataannya sangat terjadi. Oleh karena itu, penelitian ini mengusulkan pendekatan Graph Pattern Matching untuk menyelidiki pelaku anomali dalam kelompok berdasarkan tingkat usaha dan bobot atribut pelaku anomali. Studi ini menghipotesiskan bahwa tingkat usaha dan bobot atribut aktor kecurangan dapat memberikan bobot aktor penipuan untuk menemukan kepala pelaku kecurangan atau kepala mafia. Akhirnya, bobot pelaku anomali bisa digunakan untuk mengetahui apakah penyimpangan tersebut mengindikasikan, bahwa pelaku anomali atau tidak. Berdasarkan hal tersebut, pencocokan pola grafik sesuai untuk mengatasi masalah deteksi anomali pada kasus penyimpangan tinggi.

Untuk dapat mendeteksi anomali pada perusahaan PT.Terminal Peti Kemas Surabaya berdasarkan 5 Atribut tersebut yakni dengan melakukan pengumpulan data event logs dari database operational PT.Terminal Peti Kemas Surabaya. Kemudian melakukan penyeleksian data event logs yang dibutuhkan, setelah proses penyeleksian terjadi, event log disaring berdasarkan 3 faktor yakni log yang mengganggu (*Noise*), log yang tidak lengkap (*Incompleteness*) dan log terpotong (*Truncated*). Hasil log yang terfilter kemudian ditransformasi dengan ke dalam excel untuk dimodelkan ke dalam graph database yang selanjutnya akan dipakai sebagai bahan untuk mendeteksi anomali. Dari model proses yang telah digambarkan pada graph databases, proses selanjutnya adalah membentuk *Control Flow Patterns* yang digunakan untuk membuat suatu proses model yang dapat digunakan untuk mendeteksi pola yang salah dengan menggunakan *Isomorphisms Graph*. Pendeteksian anomali dilakukan secara otomatis dengan menggunakan program dan hasil deteksi anomali akan dianalisis dengan menggunakan 2 metode yaitu metode *Fuzzy Regression* dan *Multiple Linear Regression* dimana keduanya akan dibandingkan yang mana algoritma terbaik untuk mendeteksi anomali.

Multiple Linear Regression adalah algoritma untuk mengetahui tingkat prediksi anomali dari semua nilai atribut anomali sehingga nilainya *continuous* (0.1,0.2, ..., X_n). *Fuzzy Regression* adalah teknik baru untuk mengetahui tingkat prediksi kecurangan yang lebih akurat daripada regresi linier berganda, metodenya adalah mengklasifikasikan kasus mana yang memiliki anomali rendah, anomali sedang dan anomali tinggi kemudian menemukan tingkat prediksi anomali dari semua nilai atribut anomali yang nilainya *discrete* (Rendah, Sedang dan Tinggi). Hasil penelitian menunjukkan bahwa *Fuzzy Regression* lebih tepat untuk mendeteksi anomali daripada *Multiple Linear Regression*. Terbukti dengan sensitivitas dan spesifisitas *Fuzzy Regression* untuk anomali yang tinggi adalah 78% dan 99%, sensitivitas dan spesifisitas *Fuzzy Regression* untuk anomali medium adalah 85% dan 99%, sensitivitas dan spesifisitas *Fuzzy Regression* untuk anomali rendah adalah 88% dan 99% dan sensitivitas dan spesifisitas *Fuzzy Regression* untuk tidak anomali adalah 99% dan 86%.

Susunan dari thesis ini disusun sebagai berikut. Bagian 1 menyajikan latar belakang dan tujuan pembuatan sistem deteksi Anomali. Bagian 2 menyajikan ikhtisar pekerjaan terkait pada deteksi anomali. Bagian 3 menyajikan metode penelitian untuk mendeteksi tindakan dan pelaku anomali. Bagian 4 menyajikan hasil penelitian. Bagian 5 menyajikan kesimpulan.

1.2 Perumusan Masalah

Berdasarkan latar belakang diatas perumusan masalah didapatkan sebagai berikut:

1. Bagaimana mendeteksi anomali berdasarkan *Skip Sequence Analysis*, *Wrong Throughput Time Max*, *Wrong Throughput Time Min*, *Wrong Decision* dan *Wrong Pattern Analysis* dengan menggunakan algoritma *Fuzzy Regression*?
2. Bagaimana membentuk *Control Flow Patterns* untuk mendeteksi *wrong patterns* menggunakan *isomorphisms graphs*?
3. Bagaimana mengurangi *false positive* dan *false negative* pada deteksi anomali untuk meningkatkan *sensitivity* dan *specificity*?

1.3 Tujuan Penelitian

1. Membuktikan bahwa metode *Fuzzy Regression* dan *Control Flow Patterns* dapat mendeteksi anomali pada proses bisnis.
2. Mengukur dan menganalisa kinerja algoritma *Fuzzy Regression* dan *Control Flow Patterns* dalam mendeteksi anomali pada proses bisnis.

1.4 Batasan Masalah

Permasalahan pada latar belakang yang dibahas dalam penelitian ini memiliki beberapa batasan, di antaranya adalah sebagai berikut:

1. Dataset yang digunakan adalah event log kejadian nyata yang dilakukan oleh PT.Terminal Peti Kemas Surabaya yang diambil langsung dari sistem database TOS Event Log PT.Terminal Peti Kemas Surabaya.
2. Dataset Event Log terdiri dari 65000 *Case Process Discharge, Yard, Custom, Quarantine* dan *Delivery Process* (Periode Waktu: Desember s/d Maret 2016).

1.5 Kontribusi Penelitian

Pada penelitian ini memiliki kontribusi menambahkan atribut parameter pendeteksian anomali yakni deteksi anomali pada *subgraph* dari *fullgraph* proses bisnis (*Wrong Pattern Analysis*) dengan menggunakan *Control Flow Patterns* dan meningkatkan akurasi dari proses pendeteksian dengan menggunakan metode *Fuzzy Regression*.

1.6 Manfaat Penelitian

Dengan menggunakan *Fuzzy Regression* dan *Control Flow Patterns* diharapkan mampu mendeteksi anomali berdasarkan threshold nilai parameter *activity* dan *attribute* sehingga dapat dihasilkan pendeteksian anomali yang lebih optimal untuk mengatasi adanya kejadian anomali pada PT. Terminal Peti Kemas Surabaya khususnya pada kasus import.

1.7 Sistematika Penulisan

Sistematika penulisan pada penelitian ini adalah sebagai berikut:

Bab 1. Pendahuluan

Memaparkan tentang latar belakang masalah, rumusan masalah, batasan masalah, kontribusi penelitian, tujuan dan manfaat penelitian.

Bab 2. Kajian pustaka dan Dasar Teori

Menjelaskan tentang metode yang digunakan dan teori- teori pendukung dalam penelitian.

Bab 3. Metodologi Penelitian

Memaparkan dan menjelaskan tentang desain sistem serta metode yang diterapkan pada *control flow patterns* dan deteksi anomali pada event logs.

Bab 4. Uji Coba dan Analisis Hasil

Memaparkan uji coba dan menjelaskan analisis hasil kinerja algoritma.

Bab 5. Kesimpulan

Menguraikan kesimpulan yang diambil berdasarkan hasil dari penelitian yang telah diperoleh

BAB II

DASAR TEORI DAN KAJIAN PUSTAKA

Pada bab II dasar teori dan kajian pustaka akan dijelaskan untuk memberikan gambaran yang jelas yang berkaitan dengan penelitian ini. Pada bab dua dibagi menjadi empat garis besar yaitu dasar Deteksi Anomali dan Fraud, *Process Mining for Process-Based Fraud Detection*, Atribut *Process Based Fraud (PBF)*, *Data Mining* untuk Deteksi Fraud, *Association Rule Learning Applied* untuk Deteksi Fraud, *Graph Pattern Matching*, Graf Isomorfisma, *Control Flow Patterns*, *Fuzzy Regression Algorithm*.

2.1 Deteksi Anomali dan Fraud

Anomaly merupakan suatu penyimpangan atau keanehan yang sering terjadi pada suatu proses bisnis. Anomali juga sering disebut sebagai suatu kejadian yang tidak bisa diperkirakan sehingga sesuatu yang terjadi akan berubah dari kejadian biasanya. Anomali bisa menjadi suatu indikasi adanya kejadian fraud atau kecurangan dalam perusahaan.

Fraud mencakup semua bentuk kejahatan dengan tujuan mengambil keuntungan dalam berbagai cara penipuan. *Association of Certified Fraud Examiners (ACFE)* melaporkan bahwa sebuah organisasi kehilangan sekitar 5% dari pendapatan tahunan mereka. Pada tahun 2011 sendiri, fraud mencapai 3,5 triliun dollar A.S. Kerugian ini berpengaruh signifikan baik pada perusahaan kecil maupun perusahaan besar. Fraud dikelompokkan menjadi 3 jenis, yaitu *Asset Misappropriation Fraud* (Fraud terhadap aset), *Fraudulent Statements* (Fraud terhadap laporan keuangan) dan Korupsi. (Rahmawati, Yaqin and Sarno, 2016)

Penipuan bisa terjadi karena adanya pelanggaran terhadap proses bisnis / prosedur operasi standar dan manipulasi data. Untuk menganalisa konversi data, teknik data mining digunakan. Selain itu, untuk menganalisis pelanggaran dalam proses bisnis, proses penambangan digunakan. Studi deteksi kecurangan dengan menggunakan teknik data mining telah banyak diterapkan dengan menggunakan algoritma jaringan syaraf, algoritma regresi logistik dan deteksi kecurangan dengan *Web Service Collaboration*

(Chiu and Tsai, 2004). Proses berbasis Fraud (PBF) adalah penipuan yang terjadi dalam proses bisnis (Huda, Sarno and Tohari 2015). Dalam tiga tahun terakhir, jumlah Proses berbasis Penipuan (PBF) dalam total kecurangan yang ada relatif besar di Indonesia. Namun, jumlah penelitian tentang Proses Berbasis Penipuan masih kecil. Mieke Jans (Jans, Alles and Vasarhelyi, 2013) mengusulkan untuk menggunakan analisis aliran kontrol, analisis kinerja dan analisis peran untuk menganalisis log peristiwa.

Log peristiwa adalah log yang berisi informasi tentang aktivitas dalam model proses bisnis. Informasi tentang kejadian, pencetus, cap waktu dan id kasus terdapat dalam log peristiwa. Analisis flow control digunakan untuk menganalisis pola event log. Kasus yang kebugarannya hanya kecil dianggap sebagai suara bising. Kebisingan ini kemudian diidentifikasi sebagai kecurangan yang mencurigakan. Metode analisis kinerja digunakan untuk menganalisis suatu kasus yang memiliki jalur terpendek. Ini diidentifikasi sebagai kecurangan yang mencurigakan. Analisis peran digunakan untuk menganalisis kasus di mana pelanggaran pemisahan tugas terjadi. Kasus dengan beberapa peristiwa yang dieksekusi oleh pencetus yang sama, diidentifikasi sebagai kecurangan yang mencurigakan. Kendati demikian, proses pengambilan keputusan apakah itu kecurangan atau tidak, membutuhkan pendapat dari para ahli. Mieke Jans (Jans, Alles and Vasarhelyi, 2013) belum mempresentasikan penerapan teknik deteksi kecurangan dalam proses bisnis. J.Stoop (Stoop, 2012) mengusulkan Konsep 1 + 5 + 1 mengusulkan alat untuk implementasi deteksi PBF. singkatan dari "1 + 5 + 1" adalah (1) persiapan log + (5) {1} analisis log, {2} analisis proses, {3} analisis kesesuaian, {4} analisis kinerja, {5} analisis sosial menggunakan *filter*, *summarization*, *sorting*, *join* dan *aging* + (1) iterasi dan *refocusing*. Penerapan konsep tersebut adalah sebagai berikut :

- Analisis log dilakukan dengan menggunakan penyaringan, analisis *gap*, analisis duplikat dan rangkuman data.
- Analisis proses dilakukan dengan menggunakan analisis *gap*, filtering, menganalisa duplikat, penuaan, meringkas data, menganalisa statistik dan menganalisa trend..
- Analisis kesesuaian dilakukan dengan memfilter data, menganalisis *gap*, menganalisa duplikat, meringkas data dan simulasi paralel.

- Analisis kinerja dilakukan dengan menggunakan data filtering, meringkas data, stratifikasi, menyortir data, menganalisa statistik, menganalisa trend dan aging.
- Analisis sosial diimplementasikan dengan menggunakan penyaringan, meringkas data dan menyortir data.

Tabel 2.1. Penelitian Deteksi Fraud Untuk Proses Bisnis

| Peneliti | Topik Penelitian | Hasil Penelitian |
|--|--|--|
| Solichul Huda, Riyanarto Sarno, Tohari Ahmad (2016) | <i>Increasing Accuracy of Process-based Fraud Detection using a behavior model</i> | Peneliti mengeksplor dan mengkaji ulang bisnis proses pada aplikasi kredit yang mengandung tingkat penyimpangan deviasi dengan menambahkan 1 metode baru yakni <i>rating</i> PBF dengan paramater dari <i>conduite</i> sehingga terdapat 14 metode untuk deteksi fraud. |
| Solichul Huda, Riyanarto Sarno & Tohari Ahmad (2015) | Pendekatan Fuzzy MADM untuk <i>Rating Process-Based Fraud</i> | Peneliti ini telah menguraikan dan mengevaluasi proses bisnis dari aplikasi kredit application yang mengandung <i>low fraud deviations</i> . Pendekatan fuzzy digunakan untuk mendeterminasi bobot atribut PBF, sedangkan MADM digunakan untuk menentukan rating dari PBF. Metode yang digunakan sejumlah 13 metode. |

| Peneliti | Topik Penelitian | Hasil Penelitian |
|---|---|---|
| Solichul Huda, Tohari Ahmad, Riyanarto Sarno, Heru Agus Santoso (2014) | <i>Identification of process-based fraud patterns in credit application</i> | Peneliti menggunakan 10 metode deteksi fraud yakni <i>skip sequence, skip decision, wrong throughput time max, wrong throughput time min, wrong resource, wrong duty sequence, wrong duty combine, wrong duty decision, wrong pattern</i> dan <i>wrong decision</i> . |
| Riyanarto Sarno, Rahadian Dustrial Dewandono, Tohari Ahmad, Mohammad Farid Naufal and Fernandes Sinaga (2015) | <i>Hybrid Association Rule Learning and Process Mining for Fraud Detection</i> | Penelitian ini mendeteksi kecurangan dengan menganalisis penyimpangan SOP dengan menggunakan metode <i>non-fuzzy</i> . Oleh karena itu, nilai kondisinya (yaitu <i>not fraud, between not fraud and fraud, fraud, confident fraud and very confident fraud</i>) tidak ditentukan |
| D. Rahmawati, M. A. Yaqin and R. Sarno (2016) | <i>Fraud detection on event logs of goods and services procurement business process using Heuristics Miner algorithms</i> | Aplikasi untuk mendeteksi fraud menggunakan algoritma <i>heuristic miner</i> dengan menggunakan <i>threshold</i> untuk mengukur tingkat dependensi antar aktivitas. |
| B. N. A. Hidayat, A. P. Kurniati and Shaufiah (2016) | <i>Process Model Extension using Heuristics Miner</i> | Peneliti dapat menghasilkan model proses yang terbaik berdasarkan kriteria evaluasi yang sudah ditentukan (<i>precision, recall and f-measure</i>) yaitu >96%. |

| Peneliti | Topik Penelitian | Hasil Penelitian |
|--|---|--|
| S. Suriadi, M. T. Wynn, J. Xu, W. M. P. van der Aalst, and A. H. M. ter Hofstede (2017) | <i>Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs</i> | Setiap pola dijelaskan berdasarkan 8 komponen yaitu: <ul style="list-style-type: none"> • <i>Description</i> • <i>Real-life-Example</i> • <i>Affect</i> • <i>Data Quality Issues</i> • <i>Manifestation and Detection</i> • <i>Remedy</i> • <i>Side-effects of Remedy</i> • <i>Indicative Rule</i> |
| Andre B. Bondi (2000) | <i>Characteristics of scalability and their impact on performance</i> | Peneliti menghasilkan model scalable dari proses <i>network</i> dan tingkat performa bergantung pada struktur <i>scalability</i> |
| Hyeong Sik Kim, Padmashree Ravindra, Kemafor Anyanwu (2017) | <i>Type-based Semantic Optimization for Scalable RDF Graph Pattern Matching</i> | Peneliti mengembangkan teknik optimisasi pemodelan <i>scalable process</i> dan analisis menggunakan <i>graph pattern matching</i> . |

2.2 Process Mining for Process-Based Fraud Detection

Process mining adalah bidang yang muncul khusus untuk memperoleh pengetahuan dari data aktual yang tercatat dalam log peristiwa. Log peristiwa menyimpan informasi penting mengenai proses seperti jenis tugas yang dilakukan, dengan siapa tugas tertentu dilakukan, dan kapan tugas dimulai dan berakhir. Analisis informasi ini, pada gilirannya, dapat memungkinkan perusahaan melacak kembali data aktual dan kejadian yang tercatat di sistem mereka (Aalst, 2005).

Process mining menjadi dampak koneksi antara analisis proses bisnis dan data mining. *Process mining* berfokus pada analisis flow control, sedangkan data mining berkaitan dengan pengolahan data yang besar dan dilengkapi dengan analisis aliran data. Dalam kasus ini, analisis aliran kontrol dalam suatu metode, berbeda dengan aliran data yang menekankan pada inspeksi terhadap data bergerak, adalah untuk memeriksa struktur proses, tidak memindahkan data dari pada tugas dan hampir tidak memperhatikan input dan output saat tugas dilakukan (Aalst, 2010).

Dalam mempelajari *process mining*, tiga kegiatan utama, yaitu proses penemuan, pengecekan kesesuaian, dan analisis kinerja, dilibatkan. Penemuan proses mengacu pada metode untuk memperoleh model yang teramati dari algoritma log peristiwa dan penemuan (misalnya penambang heuristik, atau alfa) digunakan untuk membangun model data aktual yang teramati, yang dicatat oleh log peristiwa dan dapat diwakili dalam berbagai diagram (misalnya *Petri-Net*, *Fuzzy Model*). Pemeriksaan kesesuaian, di sisi lain, intinya adalah mengukur penyimpangan antara data riil yang tersimpan dalam event log dan model standar. Kegiatan ini dapat dilakukan dengan menggunakan algoritma (misalnya pemeriksaan kesesuaian token, atau pemeriksaan kesesuaian berbasis biaya) dan dapat mengatasi masalah terkait aktivitas yang dilewati atau disisipkan, kebisingan, atau urutan yang salah (Aalst, 2005).

Process Mining memungkinkan analisis kinerja terhadap proses. Sebagai ilustrasi, dengan menganalisis lokasi bottleneck proses, kita kemudian bisa mengatasi kemacetan dengan menambahkan lebih banyak sumber daya atau dengan menciptakan beberapa jalur alternatif proses. Tujuan dari analisis kinerja tersebut adalah untuk meningkatkan kualitas proses bisnis di perusahaan.

Dalam kasus deteksi kecurangan, *Process Mining* memberikan beberapa keuntungan. Penyesuaian konformitas bermanfaat untuk membandingkan data aktual dengan model standar karena mampu mendeteksi outlier. Nilai fitness model proses diukur untuk mengukur kesamaan, dengan sengaja mengukur seberapa dekat data aktual dibandingkan model ideal. Mengingat penyimpangan pada dua aspek adalah kebutuhan dalam pengecekan kinerja. Aspeknya adalah cara untuk memperbaiki model dan kontrol. Ini bisa mendapatkan kesesuaian yang lebih baik (Accorsi and Stocker, 2012).

Manfaat lain dari investigasi berbasis proses adalah analisis aliran kontrol. Ini bisa mendeteksi aktivitas yang dilewati, disisipkan aktivitas, dan urutan yang salah. Ini juga mengukur bagaimana log peristiwa yang diberikan sesuai dengan model proses standar. Analisis aliran kontrol memperoleh deviasi data aktual dibandingkan dengan kondisi ideal. Dalam deteksi kecurangan, bagian yang menyimpang seperti itu dianggap sebagai tipuan yang mencurigakan.

Selanjutnya, *process mining* mencakup berbagai perspektif. Ini bisa memperoleh pengetahuan dari event log dengan menekankan pada perspektif organisasi. Dalam kontrol *managerial*, beberapa prinsip dasar pengelolaan peran diterapkan, salah satunya adalah *Segregation of Duty* (SoD). Prosedur ini mewajibkan seseorang yang sah untuk melakukan setiap tugas organisasi. Untuk melakukannya, perusahaan harus memastikan bahwa karyawan yang berbeda menangani tugas yang berbeda. Berdasarkan data aktual di event log, informasi mengenai orang yang bertanggung jawab untuk melakukan tugas tertentu atau tugas yang tidak sah dapat diperoleh (Chen and Sun, 2010).

2.3 Atribut *Process Based Fraud* (PBF)

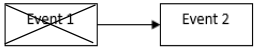
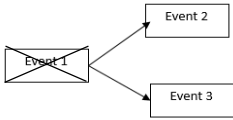
Dalam (Huda, Sarno and Tohari 2015) penulis mengajukan sepuluh atribut PBF, yaitu Analisis Salah Bentuk (*Wrong pattern analysis*), Analisis Aktivitas Skip (*Skip event analysis*), Analisis Waktu Operasi (*Throughput time analysis*), Analisis Sumber Daya (*Resource analysis*), Analisis Keputusan (*Decision analysis*), Analisis Pemisahan Tugas (*Segregation of duty analysis*), Analisis Aktivitas Bersamaan (*Parallel event analysis*), Menghitung Bobot Hubungan (*Calculation of relation weight*), Menghitung Perilaku Originator (*Calculation of originator behaviour*) dan Perhitungan nilai Bobot penyimpangan (*Calculation of Weight*

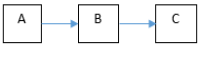

value of deviation). Namun demikian, atribut ini tidak dapat mengidentifikasi semua jenis penyimpangan dalam PBF. Berikut macam macam dari atribut PBF dan penjelasan atribut PBF tergambar pada Tabel 2.2 dan Tabel 2.3:

Tabel 2.2. Sepuluh Atribut untuk Mendeteksi Fraud (Huda, 2015)

| <i>Method Step</i> | <i>Method Analysis</i> |
|--------------------|--|
| <i>Step 1</i> | <i>Skip event analysis (Skip Sequence, Skip Decision)</i> |
| <i>Step 2</i> | <i>Throughput time analysis (Wrong Throughput Time (Min), Wrong Throughput Time (Max))</i> |
| <i>Step 3</i> | <i>Resource analysis (Wrong Resource)</i> |
| <i>Step 4</i> | <i>Decision analysis (Wrong Decision)</i> |
| <i>Step 5</i> | <i>Segregation of duty analysis (Wrong Duty Sequence, Wrong Duty Decision, Wrong Duty Combine)</i> |
| <i>Step 6</i> | <i>Wrong pattern analysis (Wrong Pattern)</i> |
| <i>Step 7</i> | <i>Parallel event analysis (Same Event execution one time)</i> |
| <i>Step 8</i> | <i>Calculation of relation weight (Value event $A \rightarrow B$)</i> |
| <i>Step 9</i> | <i>Calculation of originator behavior ('very good', 'good', 'fair', 'bad' and 'very bad')</i> |
| <i>Step 10</i> | <i>Weight value of deviation ('minor', 'medium', 'major')</i> |

Tabel 2.3. Penjelasan Atribut Deteksi Fraud

| Atribut | Penjelasan | Contoh | | | | |
|----------------------------------|---|--|----------------|---------------|----------------------------------|------------------|
| <i>Skip Sequence</i> | Jenis fraud di mana terdapat aktivitas yang di dalam SOP seharusnya dikerjakan, tetapi kenyataannya tidak dikerjakan atau dilewati pada aktivitas lurus/ <i>sequence</i> |  <p><i>Event Skip Sequence</i></p> | | | | |
| <i>Skip Decision</i> | Jenis fraud di mana terdapat aktivitas yang di dalam SOP seharusnya dikerjakan, tetapi kenyataannya tidak dikerjakan atau dilewati pada aktivitas keputusan/ <i>decision</i> |  <p><i>Event Skip Decision</i></p> | | | | |
| <i>Wrong Throughput Time Min</i> | Jenis fraud di mana terjadi pengerjaan aktivitas yang lebih cepat dari batasan waktu yang ditetapkan dalam SOP (pelanggaran waktu di bawah waktu normal) | <table border="1"><thead><tr><th>Nama Aktivitas</th><th>Waktu Standar</th></tr></thead><tbody><tr><td>Stack Container in Yard</td><td>1-2 Jam</td></tr></tbody></table> <p>Contoh: Eksekusi dari "<i>Stack Container in Yard</i>" pada log 45 Menit (lebih cepat)</p> | Nama Aktivitas | Waktu Standar | Stack Container in Yard | 1-2 Jam |
| Nama Aktivitas | Waktu Standar | | | | | |
| Stack Container in Yard | 1-2 Jam | | | | | |
| <i>Wrong Throughput Time Max</i> | Jenis fraud di mana terjadi pengerjaan aktivitas yang lebih lambat dari batasan waktu yang ditetapkan dalam SOP (pelanggaran waktu di atas waktu normal) | <table border="1"><thead><tr><th>Nama Aktivitas</th><th>Waktu Standar</th></tr></thead><tbody><tr><td>Verification Document Quarantine</td><td>45 Menit - 1 Jam</td></tr></tbody></table> <p>Contoh: Eksekusi dari "<i>Verication Document Quarantine</i>" pada log 2 Jam (lebih lambat)</p> | Nama Aktivitas | Waktu Standar | Verification Document Quarantine | 45 Menit - 1 Jam |
| Nama Aktivitas | Waktu Standar | | | | | |
| Verification Document Quarantine | 45 Menit - 1 Jam | | | | | |
| <i>Wrong Resource</i> | Jenis fraud di mana suatu aktivitas tidak dikerjakan oleh orang yang mempunyai <i>role</i> yang sesuai dengan SOP | Contoh: <i>Check Document</i> harusnya dikerjakan oleh Budi, tapi dikerjakan oleh fadli. | | | | |
| <i>Wrong Duty Sequence</i> | Jenis fraud di mana terdapat seorang pegawai yang melakukan dua atau lebih aktivitas yang berbeda dalam satu proses yang terjadi pada aktivitas <i>sequence</i> | Contoh: Proses <i>Sequence</i> " <i>Check Document</i> " dan " <i>Stack Container in Yard</i> " keduanya dikerjakan oleh orang yang sama yakni Budi, padahal harusnya dikerjakan dengan orang yang berbeda. | | | | |
| <i>Wrong Duty Decision</i> | Jenis fraud di mana terdapat seorang pegawai yang melakukan dua atau lebih aktivitas yang berbeda dalam satu proses yang terjadi pada aktivitas <i>decision</i> | Contoh: Proses <i>Decision</i> " <i>Determining Container Type</i> " dan " <i>Stack Non Reefer</i> " keduanya dikerjakan oleh orang yang sama yakni Charlie, padahal harusnya dikerjakan dengan orang yang berbeda. | | | | |
| <i>Wrong Duty Combine</i> | Jenis fraud di mana terdapat seorang pegawai yang melakukan dua atau lebih aktivitas yang berbeda dalam satu proses yang terjadi pada aktivitas <i>sequence</i> dan <i>decision</i> | Contoh: Proses <i>Sequence</i> " <i>Check Document</i> " dan " <i>Determining Container Type</i> " keduanya dikerjakan oleh orang yang sama yakni Budi, padahal harusnya dikerjakan dengan orang yang berbeda. Mengingat keduanya adalah dari jenis aktivitas yang berbeda yakni <i>sequence</i> dan <i>decision</i> . | | | | |

| Atribut | Penjelasan | Contoh |
|-----------------------|--|---|
| <i>Wrong Pattern</i> | Jenis fraud di mana terjadi kesalahan urutan aktivitas yang tidak sesuai dengan urutan aktivitas yang ditetapkan pada model proses bisnis standar di dalam SOP |   |
| <i>Wrong Decision</i> | Jenis fraud di mana terjadi kesalahan penarikan keputusan yang tidak sesuai dengan standar yang ditetapkan di dalam SOP | Contoh: Tipe kontainer tidak tahan lama (non reefer) harusnya dimasukkan ke tipe kontainer pendingin, tapi itu salah penempatan jadi ditaruh ke tipe kontainer non pendingin. |
| <i>Wrong Input</i> | Jenis fraud ini di mana terdapat kesalahan yang terjadi pada proses bisnis jika terdapat atribut dokumen <i>input</i> yang hilang atau <i>skip</i> . | Contoh: Jika ada dokumen input yang tidak ada dalam satu aktivitas maka aktivitas tersebut mempunyai anomali atau <i>wrong input</i> |
| <i>Wrong Output</i> | Jenis fraud ini di mana terdapat kesalahan yang terjadi pada proses bisnis jika terdapat atribut dokumen <i>output</i> yang hilang atau <i>skip</i> . | Contoh: Jika ada dokumen keluaran yang tidak ada dalam satu aktivitas maka aktivitas tersebut mempunyai anomali atau <i>wrong output</i> |

2.4 Data Mining untuk Deteksi Fraud

Karena pengendalian keamanan anti-penipuan memerlukan penyelidikan otomatis yang lebih kuat, penerapan metode berbasis komputer diperlukan. Banyak penelitian ditujukan untuk mengajukan metode seperti solusi otomatis untuk deteksi kecurangan. *Data mining* atau *Process Mining* disini digunakan untuk menyusun perlindungan yang kuat terhadap kasus-kasus yang tidak benar .

Data mining adalah metode komputerisasi klasik dalam analisis data besar yaitu dengan mengekstraksi abstraksi dan dengan mengolah pola data. Ini mencakup pohon keputusan, pembelajaran mesin, jaringan syaraf tiruan, atau aturan belajar asosiasi. Dalam deteksi kecurangan, dua pendekatan data mining (diawasi dan tidak diawasi) dilibatkan. Pendekatan yang diawasi memperkirakan model berdasarkan sampel transaksi palsu dan legal untuk mengkategorikan apakah transaksi baru itu legal. Dalam hal yang tidak diawasi, pencilan dikenali sebagai kecurangan yang mencurigakan. Pendekatan seperti itu memprediksi probabilitas kecurangan dalam transaksi (Ngai, Hu, Wong, Chen and Sun, 2010).

Berbagai metode data mining telah diusulkan untuk mendeteksi skema penipuan. Pohon keputusan di sini diterapkan untuk memprediksi beberapa contoh kecil yang dianggap sebagai kecurangan dan juga metode *cross*. Jaringan syaraf dan Bayesian, misalnya, diimplementasikan untuk menghapus sejumlah atribut terkait.

Support vector machine, dalam hal ini, bermanfaat untuk mencapai akurasi yang tinggi dengan data transaksi yang sangat sedikit namun tidak mampu menghadapi kuesioner baru. Dalam hal ini akan menyajikan informasi ringkas tentang kelebihan dan kekurangan metode *data mining* sebelumnya terkait dengan deteksi kecurangan. Selain metode tersebut (Ngai, Hu, Wong, Chen and Sun, 2010) mengamati bahwa wajar untuk memiliki sampel mayoritas (banyak sampel) dan minoritas (beberapa sampel) dalam deteksi kecurangan. Distribusi ketidakseimbangan ini, bagaimanapun, dapat mempengaruhi kemampuan pengklasifikasi. Jadi, disarankan untuk tidak mengabaikan kelas minoritas.

Meskipun kemampuan untuk memeriksa dataset yang besar, metode penambangan data sebenarnya tidak dapat menangani analisis berorientasi proses. Kebanyakan dari mereka adalah untuk menyelidiki pola abnormal yang ditemukan di dataset. Padahal, penipu cenderung menipu proses dengan melakukan tugas yang tidak sesuai dengan SOP. Proses pertambangan, sebagai tanggapannya, diperkenalkan sebagai jembatan antara data mining dan pemodelan proses di mana ia dapat memperoleh pemeriksaan lebih dalam mengenai proses tersebut. Secara keseluruhan, harus ada *trade-off* antara faktor kinerja dan privasi di *mining* yang harus dipertimbangkan.

2.5 Association Rule Learning Applied untuk Deteksi Fraud

Association Rule Learning (ARL) adalah salah satu metode penambangan data tanpa pengawasan dimana kumpulan item didefinisikan sebagai kumpulan satu atau beberapa item. Di sini, *support* mengacu pada rasio jumlah transaksi yang mengandung item yang ditetapkan. Keyakinan, sementara itu, berarti probabilitas bahwa kumpulan item akan ada dan diberi item lain juga ada dalam transaksi yang sama. ARL mengamati hubungan antara variabel dalam dataset. Berdasarkan perilaku yang sering ditemukan dalam dataset pelatihan, aturan asosiasi digunakan untuk deteksi dalam dataset pengujian. Dataset itu sendiri bisa menyiratkan aturan untuk berubah. Selain implementasi yang ada, aturan asosiasi juga cocok untuk menghasilkan filter melawan kecurangan. Dalam deteksi kecurangan, aturan berdasarkan data yang sering harus dipertimbangkan sehingga meningkatkan sistem untuk mendeteksi kecurangan serupa dalam kelompok berikut. (Aalst, 2010)

telah memperkenalkan implementasi ARL untuk menangkap kecurangan dalam aplikasi kartu kredit.

Dengan adanya beberapa karakteristik dan status kecurangan, aturan asosiasi diperoleh dengan menerapkan algoritma apriori. Algoritma ini menghasilkan sejumlah kandidat item set dimana beberapa kandidat kuat mendukung lebih dari ambang batas yang akan dipilih. Calon tersebut menjadi beberapa aturan asosiasi baru. C_1-C_n disini didefinisikan sebagai karakteristik X mewakili item dari karakteristik penipuan dan Y didefinisikan sebagai status penipuan. $X \Rightarrow Y$ mengacu pada aturan yang mengatakan bahwa jika item set X terjadi, Y dipertimbangkan saat itu. $\text{Supp}(X)$ menunjukkan proporsi item set X dalam dataset. Sementara itu, menunjukkan proporsi himpunan item X dan Y dalam kumpulan data. $\text{Conf}(X \Rightarrow Y)$ adalah nilai kepercayaan dari aturan $X \Rightarrow Y$.

Diperoleh dari frekuensi, saat X dan Y muncul, dibandingkan dengan frekuensi, bila hanya Y yang muncul, pada event log. Ambang batas ditetapkan untuk menentukan apakah Antecedences mempengaruhi kecurangan. Jika sebuah kasus mengandung antecedences dan tingkat kepercayaan lebih tinggi dari ambang batas, kasus ini dianggap sebagai penipuan. Sejumlah peraturan asosiasi diterapkan dalam peraturan positif dan negatif. Selain aturan $X \Rightarrow Y$, adalah mungkin untuk *mining* aturan, misalnya, $\neg X \Rightarrow Y$, $X \Rightarrow \neg Y$ atau $\neg X \Rightarrow \neg Y$. Aturan $\neg X \Rightarrow Y$, dalam kasus ini, menyiratkan bahwa jika item yang ditetapkan X tidak ada dalam database transaksi, himpunan item Y terjadi. Aturan $X \Rightarrow \neg Y$ di sisi lain menunjukkan bahwa jika item set X terjadi, Y tidak terjadi. Aturan $\neg X \Rightarrow \neg Y$ menyiratkan bahwa jika himpunan item X tidak terjadi, Y juga tidak. Kombinasi aturan seperti itu dapat digunakan sebagai aturan asosiasi negatif. Berikut merupakan perbandingan metode deteksi fraud dalam *Data Mining* yang tersaji pada Tabel 2.4:

Tabel 2.4. Metode *Data Mining* untuk Deteksi Fraud

| Reference | Method | Advantages | Disadvantages |
|------------------------------|--|---|--|
| Ehramikar (2000) | Decision tree | Predictive performance improved by increasing the number of minority instances | Only the decision tree algorithm experimented upon |
| Wheeler and Aitken (2000) | Case-based reasoning | Easiness for updating and maintaining model and robust to missing or irrelevant data | Requires two separate experiments; |
| Bolton and Hand (2001) | Outlier detection (unsupervised) | Successful in detecting local anomalies and fraudulent behavior in a continuous manner | Treating all accounts equally; not differencing between different accounts |
| Kim (2002) | Neural network with weighted fraud scores (unsupervised) | Increased number of detected frauds compared to a neural network only classifier | Back propagation used to train the neural networks; able to find local minima in the error function, optimal model may not always be reached |
| Maes (2002) | Neural & Bayesian belief networks | Improvement of fraud detection by removing highly correlated attributes | Better performance of Bayesian algorithm compared to neural networks in fraud detection |
| Chen (2004) | Support vector machine applied to questionnaire responded transaction data | Capable of achieving high accuracy in fraud detection with very little transaction data | Need to conduct new questionnaires whenever user behavior changes |
| Abdelhalim and Traore (2009) | Decision tree | Able to correctly classify 92% of the identity application fraud cases | The used data as a mix of real data collected online and synthetic data |

2.6 *Graph Pattern Matching*

Pencocokan pola grafik biasanya digunakan dalam berbagai aplikasi yang muncul seperti analisis jaringan sosial. Aplikasi ini menyoroti kebutuhan untuk mempelajari dua isu berikut. Pertama, pencocokan pola grafik secara tradisional didefinisikan dalam subgraf isomorfisma atau simulasi grafik. Namun, gagasan ini sering kali memaksakan terlalu kuat batasan topologi pada grafik untuk mengidentifikasi kecocokan yang berarti. Kedua, dalam praktiknya grafik biasanya berukuran besar, dan sering diperbarui dengan sedikit perubahan. Hal ini sering sangat mahal untuk menghitung ulang pertandingan mulai dari nol melalui algoritma batch saat grafik diperbarui (Cordella, Foggia, Sansone and Vento, 2004).

Dalam mengusulkan untuk menentukan pencocokan pola grafik berdasarkan gagasan simulasi terbatas, yang memperluas simulasi grafik dengan menentukan konektivitas simpul dalam grafik dalam jumlah hop yang telah ditentukan. Kami menunjukkan bahwa simulasi yang dibatasi dapat menemukan kecocokan yang masuk akal yang tidak dipahami oleh gagasan pencocokan tradisional. Kami juga menunjukkan bahwa pencocokan melalui simulasi dibatasi dalam waktu singkat, dengan memberikan algoritma semacam itu. Diberikan laporan hasil pada pencocokan pola grafik inkremental, untuk pencocokan yang didefinisikan dengan simulasi grafik, simulasi dibatasi, dan isomorfisma subgraf.

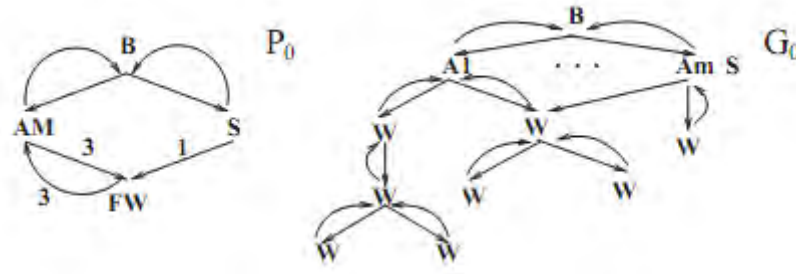
Kami menunjukkan bahwa masalah pencocokan inkremental tidak terbatas, yaitu biayanya tidak ditentukan sendiri oleh ukuran perubahan input dan output, untuk semua gagasan yang sesuai ini. Meskipun demikian, ketika pencocokan didefinisikan dalam hal simulasi atau simulasi terbatas, pencocokan inkremental semibound, itulah kompleksitas terburuknya dibatasi oleh polinomial seukuran perubahan input, output, dan informasi tambahan yang perlu diwaspadai. Menggunakan kembali perhitungan sebelumnya, dan ukuran pola grafik. pengembangan algoritma pencocokan incremental untuk simulasi grafik, dengan meminimalkan penghitungan ulang yang tidak perlu. Sebaliknya, pencocokan berdasarkan isomorfisme subgraf tidak dibatasi norsemibounded. Dengan menguji secara eksperimental keefektifan dan efisiensi algoritma ini, dan menunjukkan bahwa: (a) gagasan revisi pencocokan pola grafik memungkinkan kita mengidentifikasi komunitas yang biasanya ditemukan di jaringan kehidupan nyata, dan (b) algoritma incremental secara substansial mengungguli dalam menanggapi perubahan kecil. Ini mengindikasikan penyaranan pada kerangka kerja yang menjanjikan untuk pencocokan pola grafik kehidupan nyata.

Pencocokan pola grafik adalah untuk menemukan semua kecocokan dalam grafik data G untuk grafik pola tertentu P . Telah semakin banyak digunakan dalam penglihatan komputer, penemuan pengetahuan, biologi, cheminformatika, lalu lintas jaringan dinamis, analisis intelijen, dan analisis jaringan sosial yang baru-baru ini.

Pencocokan pola grafik biasanya didefinisikan dalam hal berikut ini.

- Subgraf isomorfisma. Ini untuk menemukan semua subgraf G yang isomorfik terhadap P (lihat Gallagher [2006] untuk sebuah survei); Yaitu, kecocokan P adalah subgraf G dari G sedemikian rupa sehingga terdapat fungsi bijektif f dari simpul P ke simpul G , dan: (a) untuk setiap simpul v pada G , v dan $f(v)$ Memiliki label yang sama, dan (b) ada tepi dari v ke v dalam P jika dan hanya jika $(f(v), f(v))$ adalah edge pada G .
- Simulasi grafik [Milner 1989]. Ini adalah untuk menemukan hubungan biner $S \subseteq VP \times V$, di mana VP dan V adalah himpunan simpul dalam P dan G , masing-masing, sehingga: (a) untuk setiap simpul u di VP , ada sebuah simpul v di V

seperti $(U, v) \in S$, dan u dan v memiliki label yang sama, dan terlebih lagi, (b) untuk masing-masing $(u, v) \in S$ dan setiap edge (u, u) pada P , ada edge (v, V) di G sedemikian rupa sehingga $(u, v) \in S$ seperti tergambar pada Gambar 2.1.



Gambar 2.1. Pattern and Data Graph

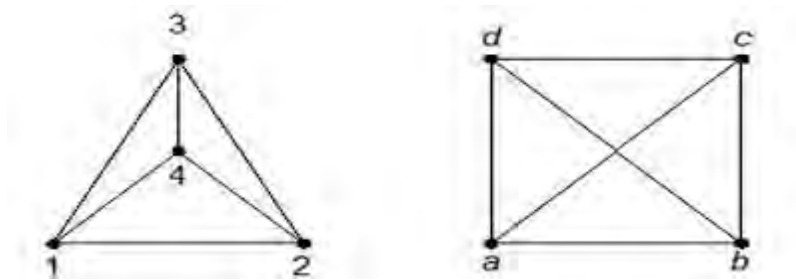
2.7 Graf Isomorfisma

Dalam geometri, dua gambar disebut kongruen jika keduanya mempunyai sifat-sifat geometri yang sama. Dengan cara yang sama, dua graf disebut isomorfis jika keduanya menunjukkan "bentuk" yang sama. Kedua graf hanya berbeda dalam hal pemberian label titik dan garisnya saja (Nabti and Seba, 2016). Secara matematis, isomorfisma 2 graf didefinisikan dalam contoh berikut :

Misalkan G adalah suatu graf dengan himpunan titik $V(G)$ dan himpunan garis $E(G)$. G' adalah graf dengan himpunan titik $V(G')$ dan himpunan garis $E(G')$. G isomorfis dengan G' bila dan hanya bila ada korespondensi satu-satu

- $V(G) \rightarrow V(G')$ dan
- $E(G) \rightarrow E(G')$

Dua buah graf, G_1 dan G_2 dikatakan isomorfik jika terdapat korespondensi satu-satu antara simpul simpul keduanya dan antara sisi-sisi keduanya. Dua buah graf yang isomorfik adalah graf yang sama, kecuali penamaan simpul dan sisinya saja yang berbeda (McKay, 1980). Ini benar karena sebuah graf dapat digambarkan dalam banyak cara seperti pada Gambar 2.2:

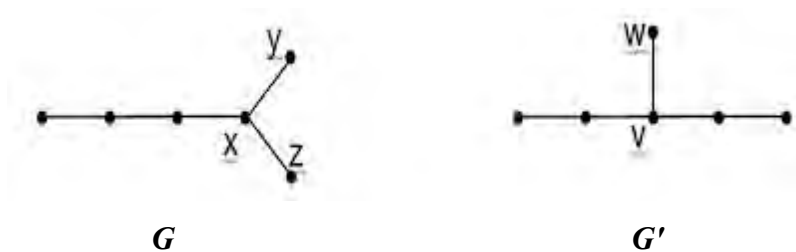


Gambar 2.2. Contoh graf isomorfis

Hingga saat ini belum ada teori yang dapat dipakai untuk menentukan apakah dua graf G dan G' isomorfis. Akan tetapi, jika G dan G' isomorfis, maka terdapat beberapa hal yang pasti dipenuhi:

- Jumlah titik G = jumlah titik G'
- Jumlah garis G = jumlah garis G'
- Jumlah garis dengan derajat tertentu dalam G dan G' sama.

Masalahnya, implikasi tersebut tidak berlaku 2 arah. Ada 2 graf yang memenuhi ketiga syarat tersebut, tetapi keduanya tidak isomorfis. Sebagai contoh adalah graf G dan G' pada Gambar 2.3 berikut ini :



Gambar 2.3. Contoh graf tidak isomorfis

Dalam G , satu-satunya titik yang berderajat 3 adalah titik x . Titik x dihubungkan dengan 2 titik lain yang berderajat 1 (titik y dan z). Sebaliknya, dalam G' , satu-satunya titik yang berderajat 3 adalah v . Satu-satunya titik berderajat 1 yang dihubungkan dengan v hanyalah titik w , sehingga G tidak mungkin isomorfis dengan G' .

2.8 *Control Flow Patterns*

Workflow Pattern dibentuk dengan tujuan untuk menggambarkan persyaratan mendasar yang timbul selama pemodelan proses bisnis secara berulang dan menggambarannya dengan cara yang imperatif. Terdapat 43 pola *Control Flow Pattern* yang telah banyak digunakan oleh praktisi, vendor dan akademisi dalam pemilihan, perancangan dan pengembangan sistem alur kerja. Tinjauan sistematis pertama dari 43 pola *Control Flow Patterns* ini memberikan deskripsi formal masing-masing dalam bentuk model *Yet Another Workflow Language* (YAWL) dan *Coloured Petri Net* (CPN). YAWL adalah bahasa pemodelan alur kerja berdasarkan apa yang disebut *Workflow Patterns*. Bahasa dan sistem pendukungnya dikembangkan dan dikelola oleh para periset di 'Eindhoven University of Technology' dan 'Queensland University of Technology'. Tujuan awal YAWL adalah mendefinisikan *Workflow Language* yang akan mendukung semua (atau sebagian besar) *Workflow Pattern* yang memiliki semantik formal. inisiatif *Workflow Pattern* bertujuan untuk membangun sebuah pendekatan yang lebih terstruktur untuk masalah spesifikasi kontrol aliran dependensi dalam bahasa alur kerja. Berdasarkan analisis sistem dan aplikasi pengelolaan *workflow* yang ada, inisiatif ini mengidentifikasi kumpulan pola yang sesuai dengan dependensi *Control Flow* yang biasa ditemukan dalam spesifikasi *workflow*, dan cara terdokumentasi untuk menangkap dependensi ini dalam *workflow language* yang ada. Pola ini bisa dijadikan tolok ukur untuk membandingkan dan mengevaluasi berbagai *workflow language* (Aalst, 2016). Tabel 2.5 mencantumkan Pola Alur Kerja Utama (*flow control perspective*).

Tabel 2.5. Control Flow Pattern

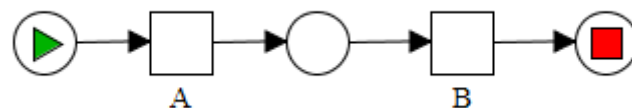
| No. | Workflow Patterns | Control Flow Pattern |
|-----|---|---|
| 1. | Basic Control Flow Patterns | <ol style="list-style-type: none"> 1. Sequence 2. Parallel Split 3. Synchronization 4. Exclusive Choice 5. Simple Merge |
| 2. | Advanced Branching and Synchronization Patterns | <ol style="list-style-type: none"> 1. Multi-Choice 2. Structured Synchronizing Merge 3. Multi-Merge 4. Structured Discriminator 5. Blocking Discriminator 6. Cancelling Discriminator 7. Structured Partial Join 8. Blocking Partial Join 9. Cancelling Partial Join 10. Generalised AND-Join 11. Local Synchronizing Merge 12. General Synchronizing Merge 13. Thread Merge 14. Thread Split |
| 3. | Multiple Instance Patterns | <ol style="list-style-type: none"> 1. Multiple Instances without Synchronization 2. Multiple Instances with a Priori Design-Time Knowledge 3. Multiple Instances with a Priori Run-Time Knowledge 4. Multiple Instances without a Priori Run-Time Knowledge 5. Static Partial Join for Multiple Instances 6. Cancelling Partial Join for Multiple Instances 7. Dynamic Partial Join for Multiple Instances |
| 4. | State-based Patterns | <ol style="list-style-type: none"> 1. Deferred Choice 2. Interleaved Parallel Routing 3. Milestone 4. Critical Section 5. Interleaved Routing |
| 5. | Cancellation and Force Completion Patterns | <ol style="list-style-type: none"> 1. Cancel Task 2. Cancel Case 3. Cancel Region 4. Cancel Multiple Instance Activity 5. Complete Multiple Instance Activity |
| 6. | Iteration Patterns | <ol style="list-style-type: none"> 1. Arbitrary Cycles 2. Structured Loop |

| No. | Workflow Patterns | Control Flow Pattern |
|-----|----------------------|--|
| | | 3. Recursion |
| 7. | Termination Patterns | 1. Implicit Termination 2. Explicit Termination |
| 8. | Trigger Patterns | 1. Transient Trigger 2. Persistent Trigger |

Dalam Control Flow Patterns (CFP) terdapat 43 model pola aliran kontrol, yang masing-masing memiliki karakteristik tersendiri, seperti pola urutan, pola paralel dan sebagainya. Dalam model proses bisnis PT. Terminal Container ada model Sequence Pattern (Gambar 2.4), Pola Split Paralel (Gambar 2.5), Pola Sinkronisasi (Gambar 2.6), pola pilihan eksklusif (Gambar 2.7) dan pola penggabungan sederhana (Gambar 2.8).

1. Sequence Patterns (Patterns 1)

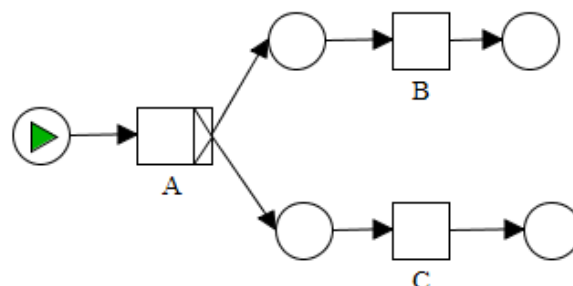
Pola 1 seperti yang terlihat pada Gambar 2.4, mendefinisikan sebuah tugas dalam proses yang diaktifkan setelah menyelesaikan tugas sebelumnya dalam proses yang sama.



Gambar 2.4. Sequence Patterns

2. Parallel Split Patterns (Patterns 2)

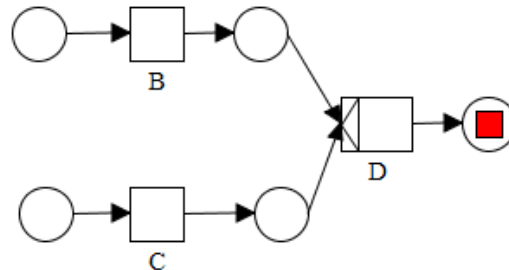
Pola 2 seperti yang terlihat pada Gambar 2.5, mendefinisikan divergensi cabang menjadi dua atau lebih cabang paralel yang masing-masing dijalankan secara bersamaan..



Gambar 2.5. Parallel Split Patterns

3. Synchronization pattern (Pattern 3)

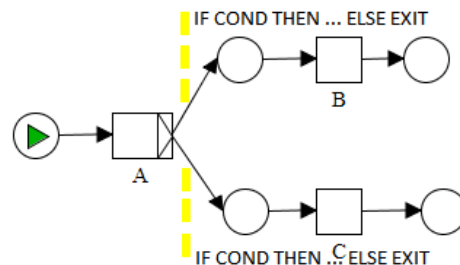
Pola 3 seperti yang terlihat pada Gambar 2.6, mendefinisikan konvergensi dua atau lebih cabang menjadi satu cabang berikutnya sehingga benang kontrol dilewatkan ke cabang berikutnya ketika semua cabang masukan telah diaktifkan.



Gambar 2.6. Synchronization pattern

4. Exclusive choice pattern (Patterns 4)

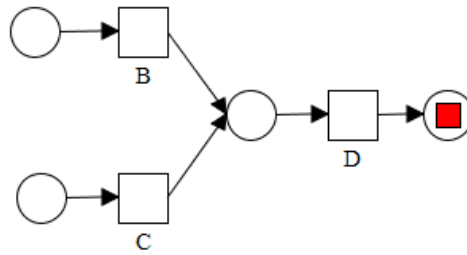
Pola 4 seperti yang terlihat pada Gambar 2.7, mendefinisikan divergensi cabang menjadi dua atau lebih cabang sehingga ketika cabang masuk diaktifkan, benang kontrol segera dilewatkan ke salah satu cabang keluar berdasarkan mekanisme yang dapat dipilih. salah satu cabang *outgoing*.



Gambar 2.7. Exclusive choice pattern

5. Simple merge pattern (Patterns 5)

Pola 5 seperti yang terlihat pada Gambar 2.8, mendefinisikan konvergensi dua atau lebih cabang menjadi satu cabang berikutnya sehingga masing-masing pemberdayaan cabang masuk menghasilkan benang kontrol yang diteruskan ke cabang berikutnya.



Gambar 2.8. Simple merge pattern

2.9 Multiple Linear Regression

Multiple linear regression merupakan algoritma untuk memodelkan hubungan antara dua atau lebih variabel penjelas dan variabel respon dengan memasukkan persamaan linier ke data yang diamati. Setiap nilai variabel independen x dikaitkan dengan nilai variabel dependen y . Garis regresi populasi p untuk variabel penjelas x_1, x_2, \dots, x_p didefinisikan sebagai $\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$. Baris ini menjelaskan bagaimana perubahan rata-rata dan perubahan dengan variabel penjelas. Nilai yang teramati untuk y bervariasi tentang meannya dan diasumsikan memiliki standar deviasi yang sama (σ). Nilai fit b_0, b_1, \dots, b_p memperkirakan parameter $\beta_0, \beta_1, \dots, \beta_p$ dari garis populasi regresi.

Karena nilai yang teramati untuk y bervariasi menurut meannya μ_y , model regresi berganda mencakup sebuah istilah untuk variasi ini. Dengan kata lain, model dinyatakan sebagai $\text{Data} = \text{Fit} + \text{Residual}$, dimana istilah "fit" mewakili ekspresi $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$. Istilah "Residual" mewakili penyimpangan nilai yang teramati y dari meannya μ_y , yang biasanya didistribusikan dengan mean 0 dan varians σ . Notasi untuk penyimpangan model adalah ε .

Secara formal, model untuk Multiple Linear Regression seperti persamaan (1) dan Gambar 2.9, yang diberikan pengamatan n adalah:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \text{ for } i = 1, 2, \dots, n. \quad (1)$$

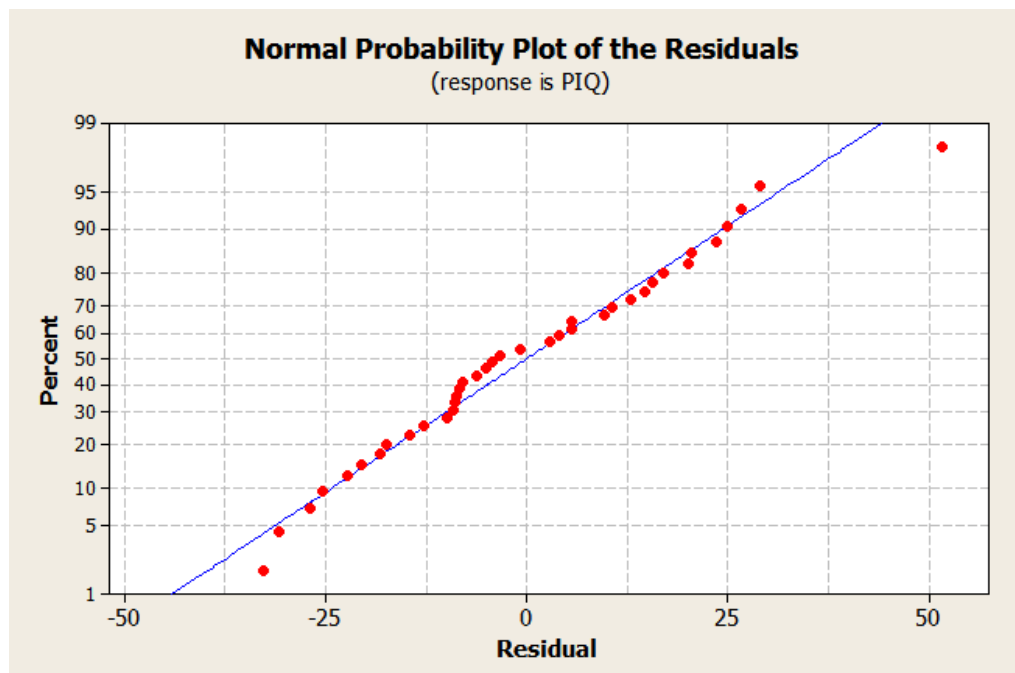
Keterangan:

y_i = Prediksi nilai variabel dependen

β_0 = Konstanta (*intercept*)

β_1 = Bobot (koefisien) regresi untuk variabel independen ke-1

x_{i1} = Variabel independen ke-1



Gambar 2.9. Model Residual Standarisasi untuk X_i - X_n .

Dalam beberapa tahun terakhir, regresi fuzzy diperkenalkan sebagai alternatif untuk mengatasi masalah klasifikasi. Pada penelitian sebelumnya, model regresi fuzzy adalah yang terbaik untuk mengklasifikasikan klasifikasi dengan menggunakan beberapa dataset (Shapiro, 2005). Hasil eksperimen tersebut menunjukkan bahwa model regresi fuzzy adalah yang terbaik untuk mengklasifikasikan suatu masalah yang terbukti dengan hasil yang lebih baik dengan menggunakan pengujian data. Analisis regresi lebih sesuai untuk sistem yang menghasilkan output kontinyu. Dalam penelitian ini, klasifikasi berbasis regresi terdiri dari satu langkah utama bahwa Multiple Linear Regression dan algoritma regresi fuzzy dievaluasi untuk menentukan tingkat prediksi suatu anomali. Untuk dataset akan digunakan dataset output diskrit karena data dari ahli tidak berkesinambungan keluaran untuk klasifikasi suatu anomali. Output regresi diterjemahkan sesuai dengan label kelas (Tidak Anomali, Anomali Rendah, Anomali Sedang dan Anomali Tinggi). Dalam penelitian ini, Multiple Linear Regression (MLR) memiliki variabel bebas x adalah *Skip Sequence*, *Wrong Throughput Time MIN*, *Throughput Time yang Salah MAX*, Keputusan Salah dan Pola yang Salah dan variabel dependen y adalah Prediksi Anomali.

2.10 Fuzzy Regression

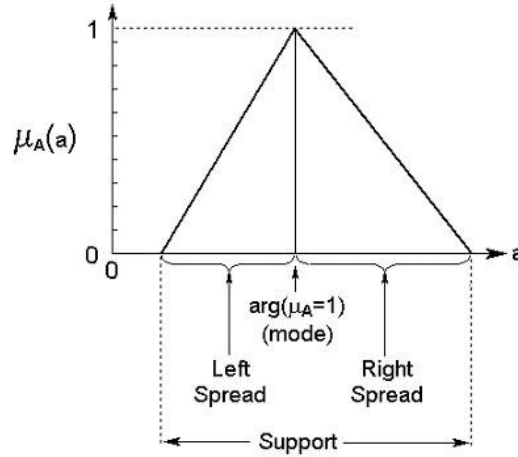
Analisis *Fuzzy Regression* memberikan hubungan fungsional fuzzy antara variabel dependen dan independen dimana ketidakjelasan hadir dalam beberapa bentuk. Data masukan mungkin tidak teratur atau kabur. Terdapat dua jenis regresi fuzzy. Yang pertama didasarkan pada konsep possibilistic dan yang kedua pada pendekatan kuadrat terkecil. Namun, dalam kedua pengertian '*best fit*' mencakup optimalisasi fungsional yang terkait dengan masalahnya. Dalam regresi *possibilistic*, fungsi ini berbentuk ukuran *spread* dari perkiraan output, baik sebagai jumlah linier tertimbang yang melibatkan koefisien estimasi dalam regresi linier, atau sebagai bentuk kuadrat dalam kasus regresi possibilistic eksponensial. Masalah optimasi ini mengurangi pemrograman linier. Untuk pendekatan kuadrat terkecil, fungsional yang harus diminimalkan adalah jarak L_2 antara keluaran yang diamati dan perkiraan. Hal ini mengurangi masalah optimasi kuadrat dan optimasi kuadrat terbatas. Metode ini dapat menggabungkan input fuzzy stokastik dan kriging fuzzy menggunakan kovarian untuk mendapatkan *Blue Estimator* (Shapiro, 2005).

Analisis regresi adalah alat analisis fundamental dalam banyak disiplin ilmu. Metode ini menganalisis hubungan antara variabel dependen dan independen berdasarkan data yang diberikan dari sudut pandang statistik, yaitu data pengamatan acak dengan beberapa kesalahan pengukuran atau noise. Di sisi lain, hubungan antara variabel dependen dan independen dapat ditandai dengan hubungan fungsional fuzzy untuk data tertentu yang mungkin kabur atau garing. Artinya, fenomena yang tidak pasti harus dimodelkan oleh hubungan fungsional fuzzy. Berdasarkan gagasan ini, Tanaka dkk. mengusulkan model regresi probabilitas dimana sistem linear fuzzy digunakan sebagai model regresi (Tanaka, Uejima and Asai, 1982). *Fuzzy Regression Model* dinyatakan sebagai:

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 X_1 + \cdots + \tilde{A}_n X_n \quad (2)$$

Dimana \tilde{Y} untuk Fuzzy Output, A_i adalah koefisien fuzzy dan X_n adalah *n-dimensional non fuzzy input vector*. Komponen fuzzy diasumsikan sebagai *Triangular Fuzzy Number* (TFNs). Konsekuensinya, untuk koefisien, contohnya,

dapat dikarakteristikan sebagai *Membership Function* (MF), $\mu_A(a)$, adalah representasi dari Gambar 2.10 berikut:



Gambar 2.10. Koefisien Fuzzy

Seperti yang ditunjukkan pada Gambar 2.5, fitur menonjol dari TFN adalah mode, penyebaran kiri dan kanannya, dan *support*-nya. Bila kedua *spread*-nya sama, TFN dikenal sebagai TFN simetris (STFN).

Ide dasar pendekatan Tanaka, yang sering disebut sebagai regresi possibilistic, adalah meminimalkan ketidakjelasan model dengan meminimalkan penyebaran total koefisien fuzzy, yang dapat mencakup semua data yang diberikan.

Berikut akan dijelaskan dasar-dasar Regresi Linier Fuzzy mencakup motivasi untuk FR (Fuzzy Regresi), komponen FR, koefisien fuzzy, faktor h-tertentu, dan output fuzzy.

1. Dasar Fuzzy Regresi

Bentuk dari Regresi linier statistik klasik :

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i, i = 1, 2, \dots, m \quad (3)$$

Dimana variabel dependen (response), y_i , variabel independen (explanatory), x_{ik} , dan koefisien (parameter), β_k , adalah nilai crisp, dan ε_i adalah istilah error acak rangkap dengan $E(\varepsilon_i) = 0$, varians $\sigma^2(\varepsilon_i) = \sigma^2$, dan kovariansi $\sigma(\varepsilon_i, \varepsilon_j) = 0, \forall i, j, i \neq j$.

Meskipun regresi statistik memiliki banyak aplikasi, masalah dapat terjadi pada situasi berikut:

- Jumlah pengamatan tidak memadai (kumpulan data kecil)

- Kesulitan untuk memverifikasi asumsi distribusi
- Ketidakjelasan dalam hubungan antara variabel input dan output
- Ambiguitas kejadian atau tingkat kejadiannya
- Ketidaktepatan dan distorsi diperkenalkan dengan linearisasi

Jadi, regresi statistik bermasalah jika kumpulan data terlalu kecil, atau ada kesulitan untuk memverifikasi bahwa kesalahan terdistribusi normal, atau jika ada ketidakjelasan dalam hubungan antara variabel independen dan dependen, atau jika ada ambiguitas yang terkait dengan kejadian tersebut. atau jika asumsi linearitas itu tidak tepat. Ini adalah situasi yang sangat tepat untuk mengusulkan regresi fuzzy agar masalah diatas dapat diatasi.

2. Komponen Fuzzy Regresi

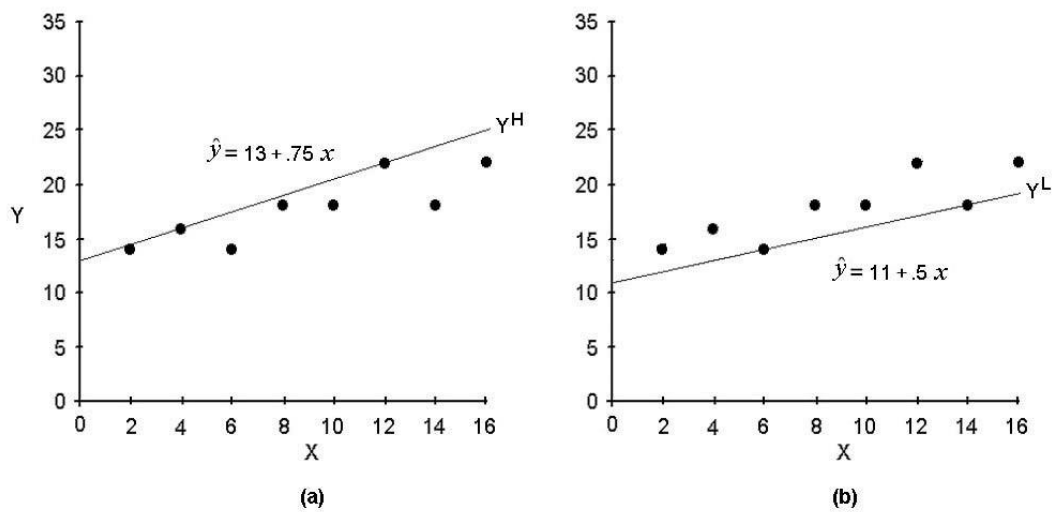
Ada dua cara umum (tidak harus saling eksklusif) untuk mengembangkan model regresi fuzzy: (1) model dimana hubungan variabel tidak jelas; dan (2) model dimana variabel itu sendiri tidak jelas. Kedua model ini dieksplorasi dalam sisa artikel ini, namun, untuk konseptualisasi ini, kami berfokus pada model dimana data renyah dan hubungan variabelnya tidak jelas.

Ini adalah masalah sederhana untuk mengkonseptualisasikan regresi fuzzy. Contoh data Ishibuchi (1992) sederhana seperti Tabel 2.6 berikut ini.

Tabel 2.6. Data Pairs

| | | | | | | | | |
|-------|----|----|----|----|----|----|----|----|
| i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| x_i | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 |
| y_i | 14 | 16 | 14 | 18 | 18 | 22 | 18 | 22 |

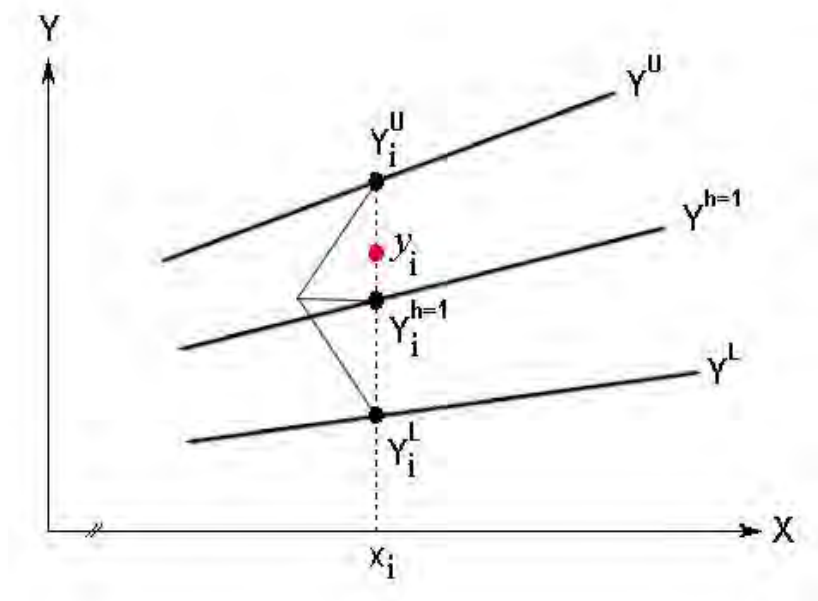
Dimulai dengan data ini, cocokkan dengan garis lurus melalui dua atau lebih titik data sedemikian rupa sehingga batas poin data dari atas. Di sini, titik-titik ini ditentukan secara heuristik dan OLS (*Ordinary Least Squares*) digunakan untuk menghitung parameter garis berlabel Y^H , yang mana mengambil $\hat{y} = 13 + .75x$, seperti yang ditunjukkan pada Gambar 2.11(a) berikut:



Gambar 2.11 Konseptualisasi upper dan lower bound

Demikian pula, kita cocokkan dengan garis lurus kedua melalui dua atau lebih titik data sedemikian rupa sehingga membatasi titik data dari bawah. Seperti ditunjukkan pada Gambar 2.11 (b), garis pas di Kasus ini diberi label Y^L dan mengambil nilai $\hat{y} = 11 + .5x$.

Dengan asumsi, untuk tujuan contoh ini, STFNN digunakan untuk MFs, mode MF jatuh di tengah jalur batas. Untuk setiap pasangan data tertentu, (x_i, y_i) , konseptualisasi di atas dapat diringkas dengan interval regresi fuzzy $[Y^L, Y^U]$ yang ditunjukkan pada Gambar 2.12.



Gambar 2.12. Interval Fuzzy Regresi

$Y_i^{h=1}$ adalah mode untuk MF dan jika SFTN diasumsikan, $Y_i^{h=1} = \bar{Y}_i = (Y_i^U + Y_i^L)/2$. Diberikan parameter, $(Y_i^U, Y_i^L, Y_i^{h=1})$, yang mengkarakterisasi model regresi fuzzy, *i-th*, pasangan data (x_i, y_i) , dikaitkan dengan parameter model $(Y_i^U, Y_i^L, Y_i^{h=1})$. Dari persamaan regresi, kita dapat melihat $Y_i^U - y_i$ dan $y_i - Y_i^L$ sebagai komponen SST, $y_i - Y_i^{h=1}$ sebagai komponen SSE, dan $Y_i^U - Y_i^{h=1}$ dan $Y_i^{h=1} - Y_i^L$ sebagai komponen SSR, yang dibahas oleh Wang dan Tsaur (2000).

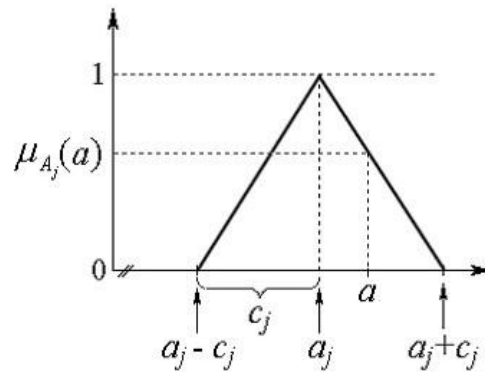
Dalam analisis probilistik berdasarkan STFN, hanya titik data yang terlibat dalam menentukan batas atas dan bawah penentuan struktur model, seperti yang digambarkan pada Gambar 2.13. Sisa titik data tidak berdampak pada struktur. Masalah ini diatasi dengan menggunakan TFN asimetris.

3. Koefisien Fuzzy

Menggabungkan Persamaan (3) dan Gambar 2.12, dan untuk saat ini, yang membatasi pembahasan pada STFN, MF dari koefisien ke- j , dapat didefinisikan sebagai:

$$\mu_{A_j}(a) = \max\left\{1 - \frac{|a - a_j|}{c_j}, 0\right\} \quad (4)$$

Dimana a_j adalah mode dan c_j adalah spread, dan ditunjukkan seperti pada Gambar 2.13:



Gambar 2.13. Parameter fuzzy simetris

Mendefinisikan:

$$\tilde{A}_j = \{a_j, c_j\}_L = \{\tilde{A}_j: a_j - c_j \leq \tilde{A}_j \leq a_j + c_j\}_L, j = 0, 1, \dots, n \quad (5)$$

Dan membatasi pertimbangan pada kasus dimana hanya koefisien yang tidak jelas, kita dapat menulis:

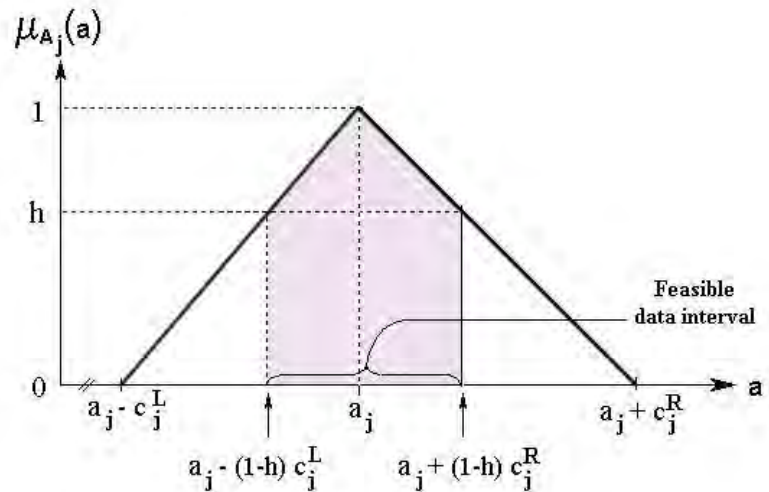
$$\tilde{Y}_i = \tilde{A}_0 + \sum_{j=1}^n \tilde{A}_j x_{ij} = (a_0, c_0)_L + \sum_{j=1}^n (a_j c_j)_L x_{ij} \quad (6)$$

Ini adalah formulasi yang berguna karena secara eksplisit menggambarkan modus dan penyebaran parameter fuzzy.

4. Faktor "*h-certain*"

Jika, seperti pada Gambar 2.14, dukungan hanya cukup untuk memasukkan semua titik data sampel, hanya ada kepercayaan terbatas pada proyeksi di luar sampel dengan menggunakan model FR yang diperkirakan. Ini dipecahkan untuk FR, sama seperti dengan regresi statistik, dengan memperluas *support*.

Perhatikan MF yang terkait dengan koefisien fuzzy *j-th*, sebuah representasi yang ditunjukkan pada Gambar 2.14.

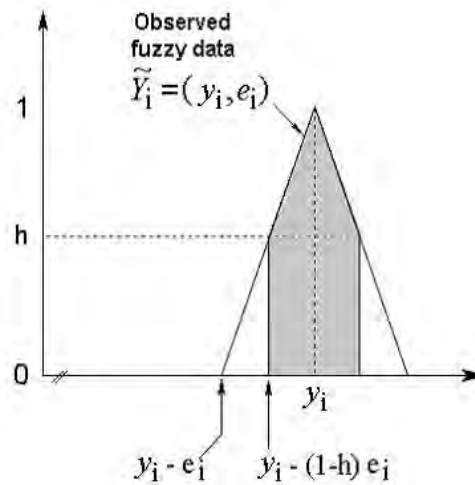


Gambar 2.14. Estimasi A_j menggunakan faktor "*h-certain*"

Untuk tujuan ilustrasi, TFN non-simetris ditampilkan, di mana c^L dan c^R mewakili penyebaran kiri dan kanan masing-masing. Di luar itu, yang membuat MF berbeda dari yang ditunjukkan pada Gambar 2.15, adalah bahwa ia mengandung titik "*h*" pada sumbu y, yang disebut faktor "*h-certain*" yang, dengan mengendalikan ukuran yang layak antar interval data (basis area yang diarsir), memperluas dukungan MF. Secara khusus, Jika faktor *h* meningkat untuk kumpulan data tertentu, maka meningkatkan *spread*, c^L dan c^R .

5. Pengamatan Keluaran Fuzzy

Faktor “*h-certain*” tertentu juga dapat diterapkan pada keluaran fuzzy yang diamati. Dengan demikian, data keluaran ke-*i* mungkin diwakili oleh STF_N, $\tilde{Y}_i = (y_i, e_i)$, di mana y_i adalah mode dan e_i adalah penyebarannya, seperti ditunjukkan pada Gambar 2.15. Di sini, titik data sebenarnya berada dalam interval $y_i \pm (1-h) e_i$, dasar dari bagian yang diarsir dari grafik.



Gambar 2.15. Pengamatan Keluaran Fuzzy

6. Fitting Model Regresi Fuzzy

Dua pendekatan umum digunakan agar sesuai dengan model regresi fuzzy:

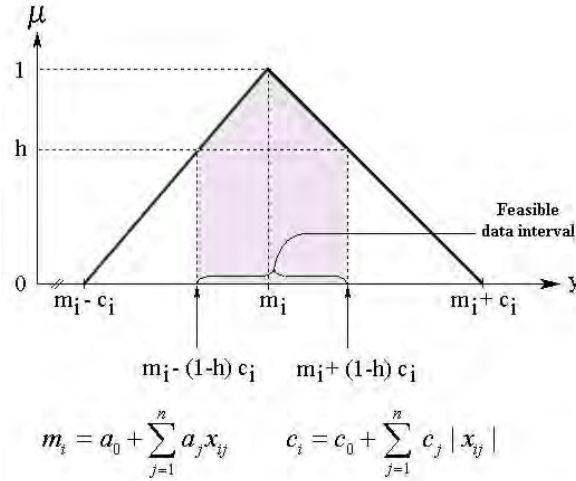
1. Model Possibilistik. Meminimalisasi ketidakjelasan model dengan meminimalkan spread total koefisien fuzzy-nya (lihat Gambar 2.19), dengan subjek memasukkan titik data setiap sampel dalam interval data yang layak.
2. Model kuadrat terkecil. Minimalkan jarak antara output model dan keluaran yang diamati, berdasarkan mode dan spreadnya.

7. Model Regresi Possibilistik

Model regresi possibilistic dioptimalkan dengan meminimalkan penyebaran, tergantung pada penahanan data yang memadai, penyebarannya diminimalkan.

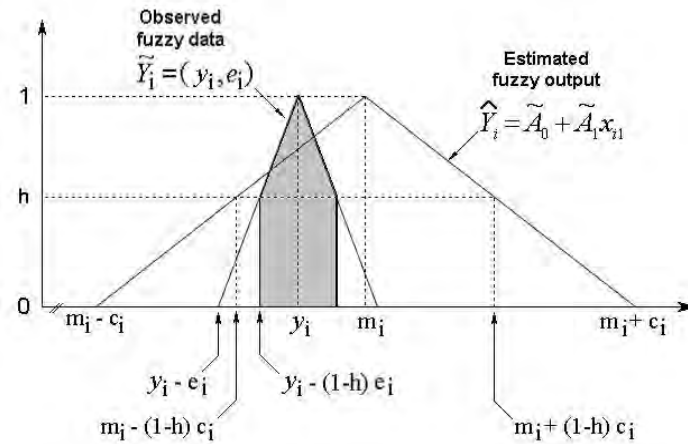
$$\min[c_0 + \sum_{j=1}^n c_j |X_{ij}|], c_j \geq 0 \quad (7)$$

Gambar 2.16 menunjukkan langkah pertama dalam persyaratan penahanan, dengan menunjukkan bagaimana Gambar 2.16 dapat dengan mudah diperluas untuk menggambarkan output fuzzy dari model.



Gambar 2.16. Keluaran Fuzzy dari Model

Gambar 2.17 berikut adalah representasi bagaimana perkiraan keluaran fuzzy dapat disesuaikan dengan data fuzzy yang diamati.



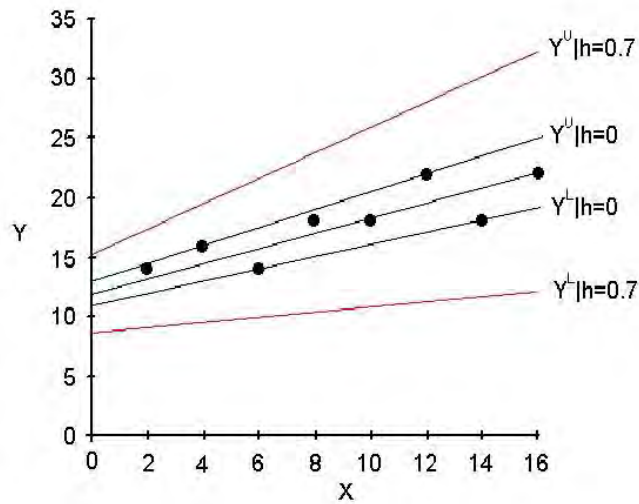
Gambar 2.17. Fitting perkiraan output ke output yang diamati

Kuncinya adalah bahwa data fuzzy yang diamati, yang disesuaikan dengan faktor “*h-certain*”, terkandung dalam perkiraan keluaran fuzzy, disesuaikan dengan faktor “*h-certain*”. Secara formal, Gambar 2.18 menunjukkan dampak faktor-*h* pada data sampel, diberikan $h = 0$ dan $h = .7$.

$$a_0 + \sum_{j=1}^n a_j x_{ij} + (1-h)[c_0 + \sum_{j=1}^n c_j |x_{ij}|] > y_i + (1-h)e_i \quad (8)$$

$$a_0 + \sum_{j=1}^n a_j x_{ij} - (1-h)[c_0 + \sum_{j=1}^n c_j |x_{ij}|] < y_i - (1-h)e_i$$

$$c_j \geq 0, i = 0, 1, \dots, m, j = 0, 1, \dots, n$$



Gambar 2.18. Model Fuzzy Linear Regression dan model h -certain

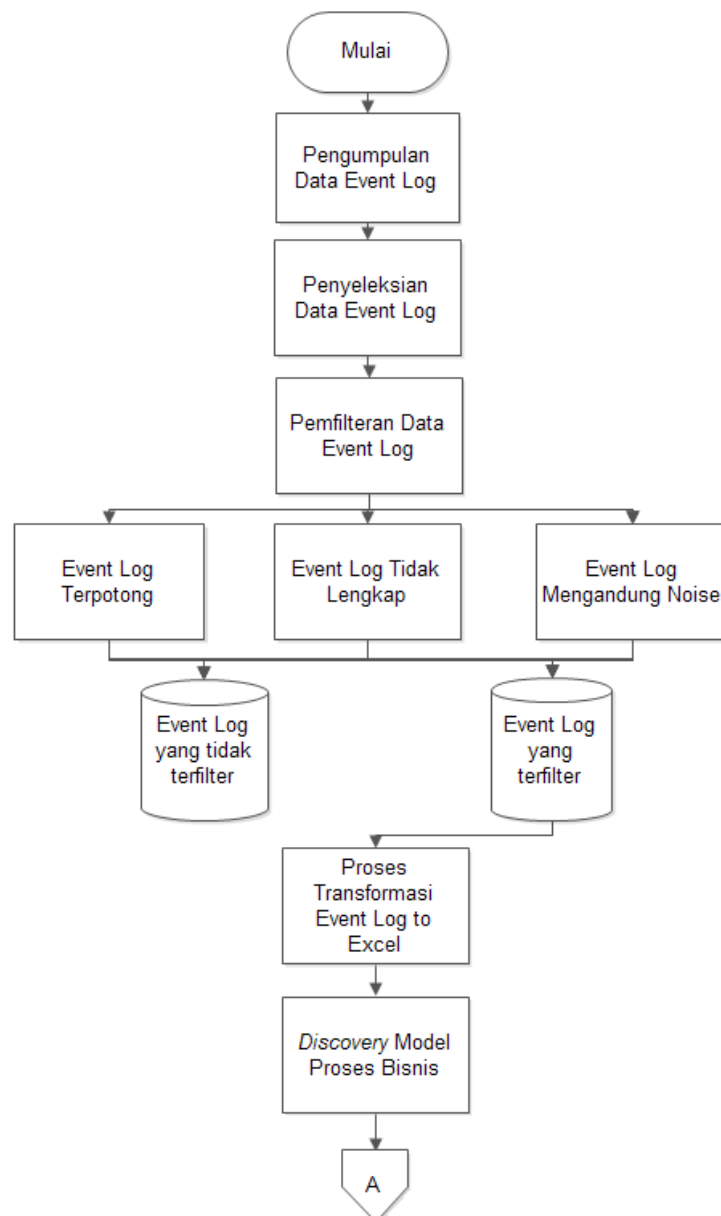
Hasil dari *Fuzzy Regression* ini adalah sesuai dengan apa yang diharapkan. Meningkatkan faktor- h , memperluas interval kepercayaan dan dengan demikian, meningkatkan probabilitas dengan nilai di luar sampel akan masuk dalam model. Hal ini sebanding dengan meningkatnya kepercayaan pada regresi statistik dengan meningkatkan interval kepercayaan.

[Halaman ini sengaja dikosongkan]

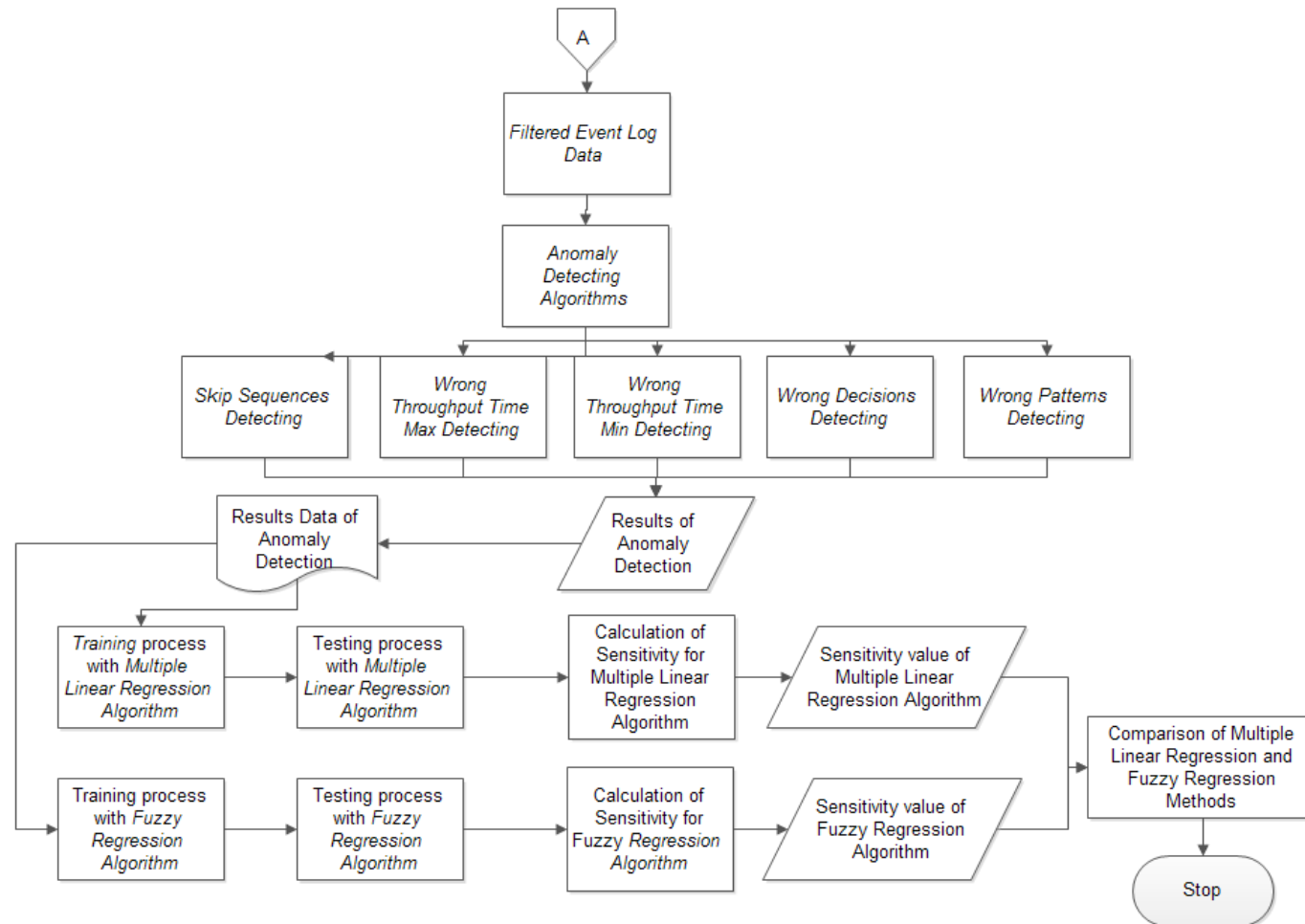
BAB III

METODE PENELITIAN

Pada metode penelitian menjelaskan desain sistem yang akan dibangun di dalam penyusunan tesis ini. Desain sistem secara umum mengenai kontribusi penelitian deteksi fraud dapat dilihat pada Gambar 3.1 dan 3.2 berikut:



Gambar 3.1 Desain Sistem



Gambar 3.2. Desain Sistem (Lanjutan)

Pada Desain sistem Gambar 3.1 dan Gambar 3.2 dimulai dengan pengumpulan data event log dari PT.Terminal Peti Kemas Surabaya, Event log dilakukan penyeleksian dengan menghapus data log yang tidak diperlukan, dari semula terdapat 41 kolom label dan 65000 data dari event log menjadi 9 kolom label dan 50 data yang diperlukan untuk dideteksi. Kemudian dari data penyeleksian log yang dilakukan, proses selanjutnya adalah memfilter data log dengan menggunakan 4 strategi event log yakni pemfilteran data log yang terpotong (*truncated*), data log yang tidak termasuk dalam proses bisnis (*noise*), data log yang terjadi bersamaan (*overlap log*) dan data log yang atributnya hilang (*missing attribute*). Pada tahap selanjutnya data event log akan melakukan setelah proses penyeleksian terjadi, event log di saring (*filter*) berdasarkan 3 faktor yakni log yang menyimpang (*Anomaly*), log yang tidak lengkap (*Incompleteness*) dan log terpotong (*Truncated*). Hasil log yang terfilter kemudian ditransform ke dalam excel untuk dimodelkan ke dalam graph databases yang selanjutnya akan dipakai sebagai bahan untuk mendeteksi anomali atau fraud. Sedangkan Hasil log yang tidak terfilter (log yang mengandung *Anomaly*, *Incompleteness* dan *Truncated*) kemudian disimpan dan diperbaiki hingga menghasilkan log yang sesuai/benar dan tidak mengandung 3 faktor terjadinya *filter* sehingga kemudian bisa dimasukkan ke dalam hasil log yang terfilter.

Dari model proses yang telah digambarkan pada graph databases, proses selanjutnya adalah mendeteksi anomali dengan menggunakan 6 parameter atau atribut secara otomatis dengan menggunakan program dan hasil deteksi anomali tersebut nantinya akan dianalisis dengan menggunakan 2 metode, analisis dengan menggunakan metode Fuzzy Regression dan analisis dengan menggunakan Multiple Linear Regression.

Untuk menguji kebenaran hasil Fuzzy Regression dan Multiple Linear Regression dengan menggunakan confusion matrix dan penghitungan nilai sensitivity dan specificity dari pendeteksian anomali pada setiap case proses bisnis. Diharapkan penelitian ini menghasilkan metode pendeteksian yang lebih akurat dan

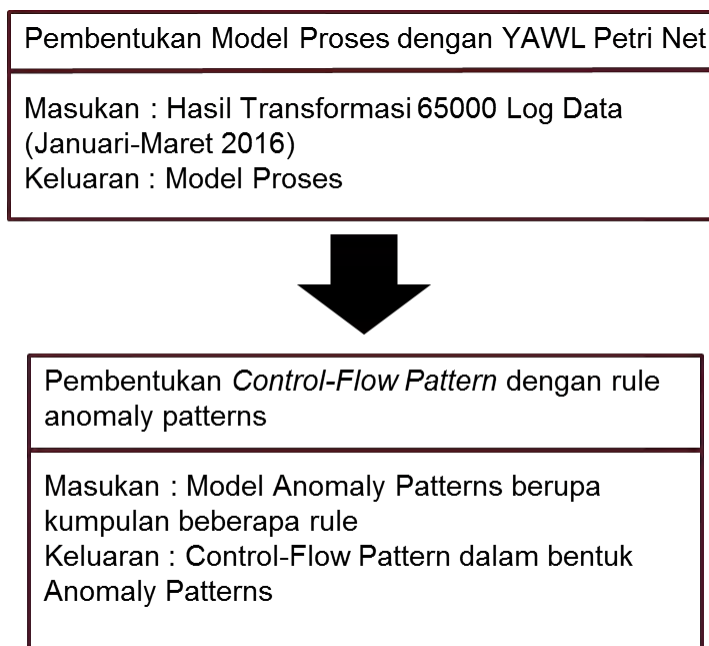
dapat menurunkan nilai False Positive dan nilai False Negative pada Kasus Anomali.

3.1 Sumber Data Penelitian

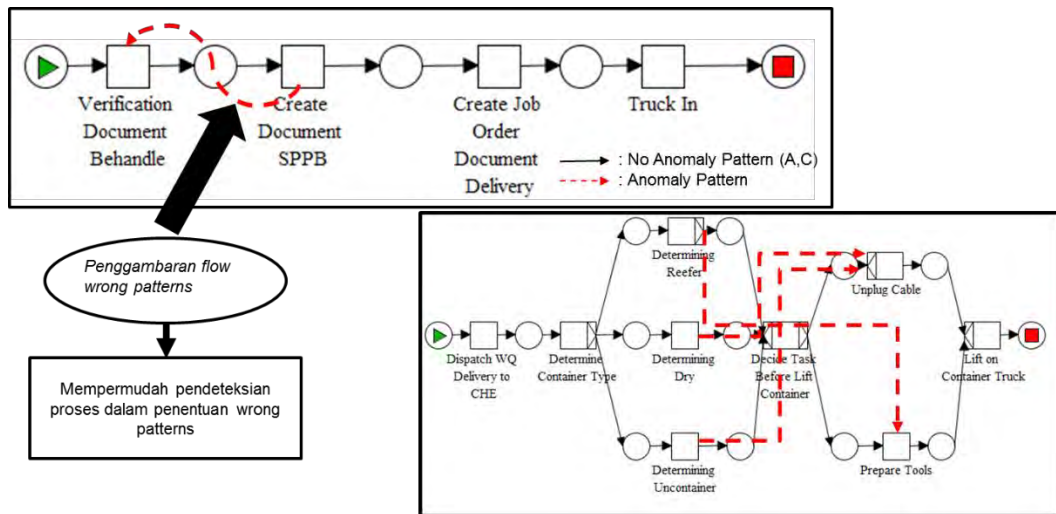
Pengumpulan data penelitian dilakukan di PT. Terminal Peti Kemas Surabaya dengan mengambil event log pada database sistem perusahaan yang kemudian dilakukan penyeleksian dengan menghapus data log yang tidak diperlukan, dari semula terdapat 62 kolom label dan 65000 data dari event log menjadi 12 kolom label dan 60% data sebagai data training dan 40% case data sebagai testing. Gambar 3.4 merupakan gambaran proses yang terjadi dalam terminal petikemas, Tabel 3.6 merupakan database sebelum ditransformasi dan Tabel 3.7 merupakan database setelah ditransformasi.

3.2 Pembentukan *Control-Flow Pattern* dengan rule *anomaly patterns*

Metode yang digunakan untuk pembentukan Control-Flow Pattern dengan rule anomaly patterns adalah dengan membentuk flow control anomaly yang mana outputnya akan berbentuk petri net dari anomaly patterns. Berikut bagan gambaran proses input output pembentukan anomaly patterns seperti pada Gambar 3.3.



Gambar 3.3. Bagan gambaran proses input output pembentukan anomaly patterns



Gambar 3.4. Build Control Flow Patterns

Pembentukan Control Flow Pattern dimana rules yang digunakan adalah No Anomaly Pattern dan Anomaly Pattern. Gambar 3.4 merupakan cara pembentukan *Control Flow Patterns* dan Gambar 3.5 merupakan Kombinasi dari rules yang digunakan dalam penentuan Control Flow Patterns.

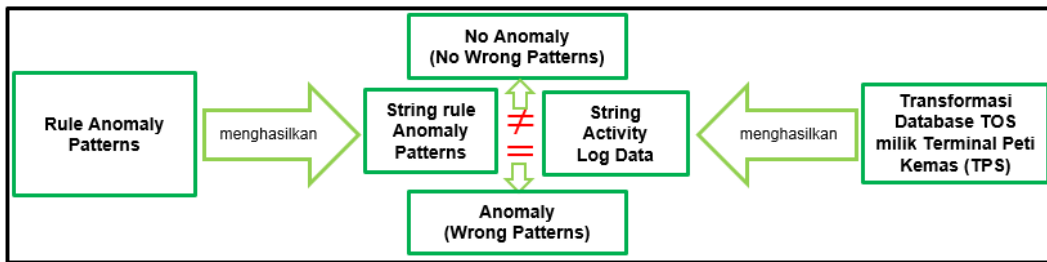
| Rules Anomaly | Deskripsi |
|--------------------------|---|
| No Anomaly Pattern (A,C) | Jika A terekam, maka aktivitas yang langsung terekam selanjutnya adalah C (Tidak Anomaly) |
| Anomaly Pattern (A,C) | Jika C terekam, maka aktivitas yang langsung terekam selanjutnya adalah A (Anomaly) |

Gambar 3.5. Rules Anomaly Pattern

3.3 Pendeteksian Wrong Patterns dengan Rule Anomaly Patterns

Setelah control-flow pattern terbentuk, maka control-flow tersebut disusun menjadi model proses dalam bentuk model anomaly pattern (seperti alur pada gambar 3.6). Langkah dalam pembentukan model anomaly pattern adalah:

1. Membangun model dari control-flow pattern.
2. Control-flow pattern dalam bentuk rule kemudian diubah menjadi sebuah string rule di program java untuk digunakan sebagai pendeteksian anomaly wrong patterns di dalam event log.



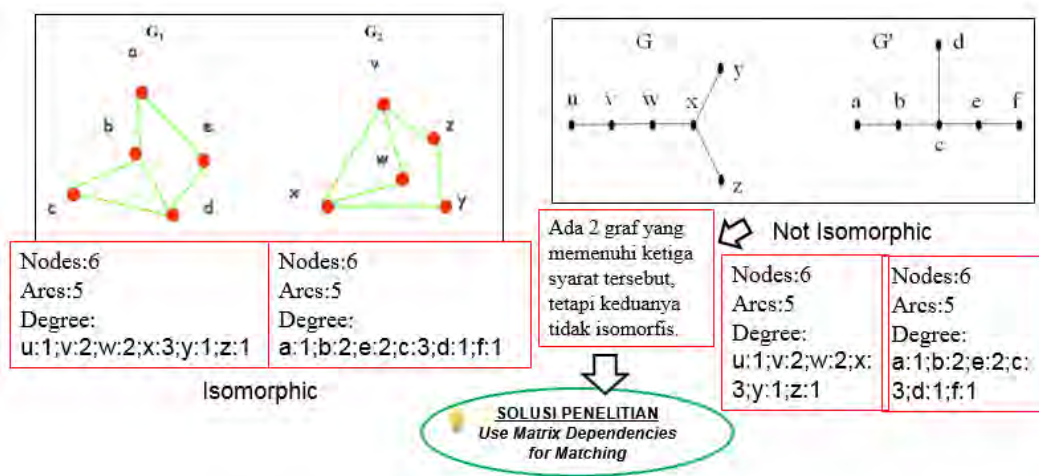
Gambar 3.6. Alur Pendeteksian Wrong Patterns dengan Rule Anomaly Patterns

Setelah mendapatkan string rule dengan program, kemudian menghitung *number of nodes*, *number of arcs*, *degree of nodes* yang kemudian digunakan sebagai parameter untuk melakukan pendeteksian anomali wrong pattern menggunakan *Isomorphisms Graph (Graph Pattern Matching)*. Dikatakan Anomaly Wrong pattern jika Graph tersebut Isomorphic atau *graph pattern matching* dengan anomaly pattern.

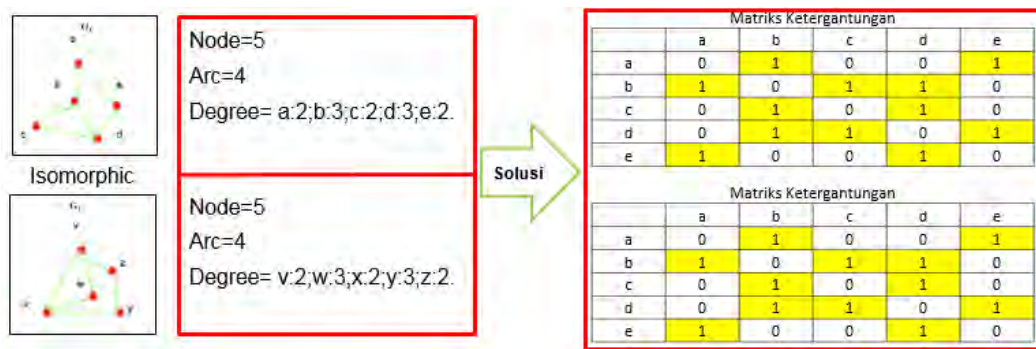
Faktor Pendukung *Graph Isomorphic*:

1. *Has the same number of nodes.*
2. *Has the same number of arcs.*
3. *Has the same degree of nodes.*

Namun meskipun sudah memenuhi ketiga syarat tersebut, tapi keduanya tidak *isomorphic*, untuk itu dibutuhkan faktor pendukung lain yakni *dependencies matrix* untuk mendeteksi kesamaan graph atau pola atau yang disebut *Graph Pattern Matching* seperti pada Gambar 3.7, Gambar 3.8 dan Gambar 3.9.

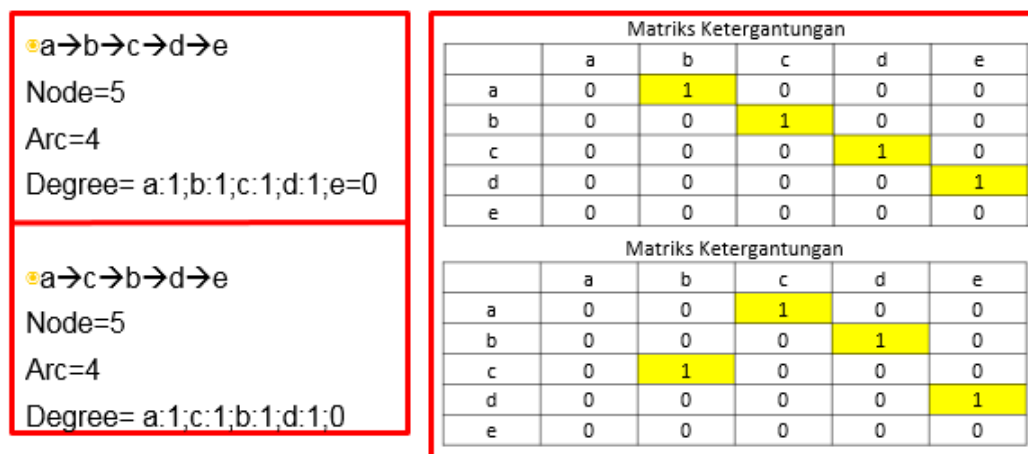


Gambar 3.7. Contoh Penerapan Isomorphisms Graph pada Case



Graf Isomorfisme, karena jumlah node, arc dan degree sama dan matriks ketergantungan sama
TERDETEKSI WRONG PATTERN

Gambar 3.8. Contoh Penerapan *Isomorphisms Graph with dependencies matrix* pada Case mengandung anomali



Graf Isomorfisme, karena jumlah node, arc dan degree sama tapi matriks ketergantungan tidak sama
TIDAK TERDETEKSI WRONG PATTERN

Gambar 3.9. Contoh Penerapan *Isomorphisms Graph with dependencies matrix* pada Case tidak mengandung anomali

3.4 Pendeteksian Anomali dengan Program Java

Untuk mendeteksi anomali secara otomatis pada banyak log di TPS Surabaya, dalam penelitian ini dibuat sebuah program berbasis JAVA yang inputnya dari excel kemudian diimpor ke SQL kemudian mendeteksi dengan menggunakan program yang telah dibuat dengan logika seperti Pseudocode pada Gambar 3.10, 3.11, 3.12 dan 3.13 berikut:

```

1. for i = 0 to activity in one case_SOP
2.   if activity_sop not same as activity_log
3.     && not decision_activity
4.       skip_activity ++
5.       If decision_activity and not same

```

Gambar 3.10. Pseudocode for Skip Sequences and Skip Decision

```

1. for i = 0 to activity in one case_SOP
2.   if time_cost_activity_log lower than
3.     standard_time
4.       throughputMin++
5.   if time_cost_activity_log higher than

```

Gambar 3.11. Pseudocode for Throughput Time Min/Max

```

1. for i = 0 to activity in one case_SOP
2.   if activity_sop not same as activity_log
3.     patternSop[] <- activity_sop
4.     patternLog[] <- activity_log
5.   if patternSop not same as patternLog
6.     wrongPattern += total_wrong_index
7.   return wrongPattern / patternLog size

```

Gambar 3.12. Pseudocode for the Wrong Pattern

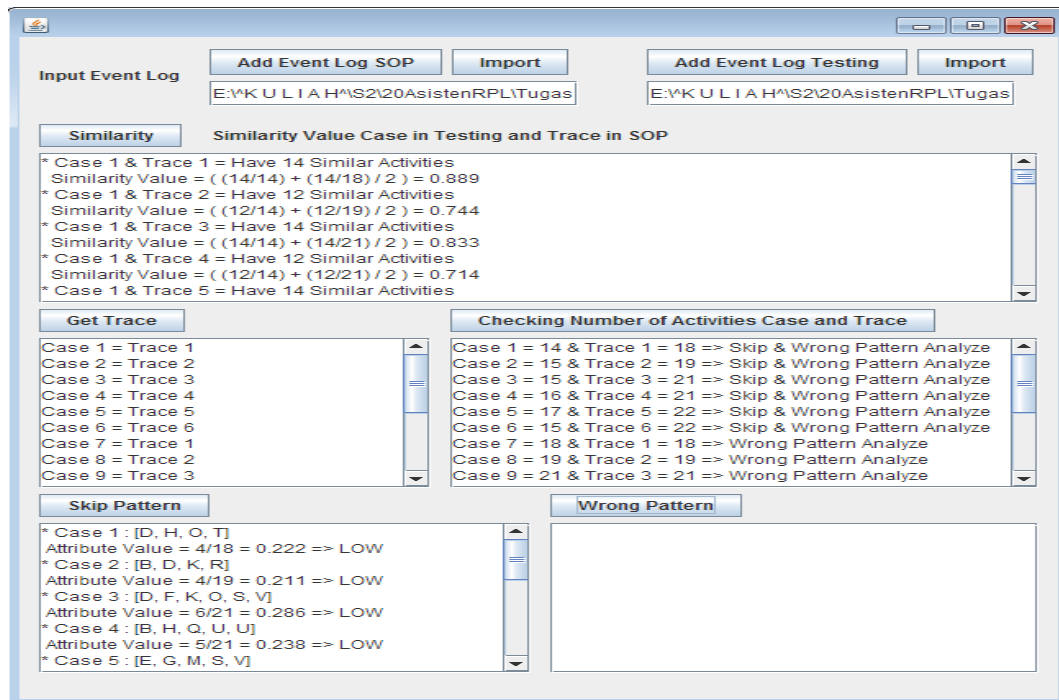
```

1. for i = 0 to activity in one case_SOP
2.   if type_container is dry
3.     if yard_block_log not same as SOP
5.       wrongDecision++
6.     if yard_slot not same as SOP
7.       wrongDecision ++
8.   return wrongDecision / 2

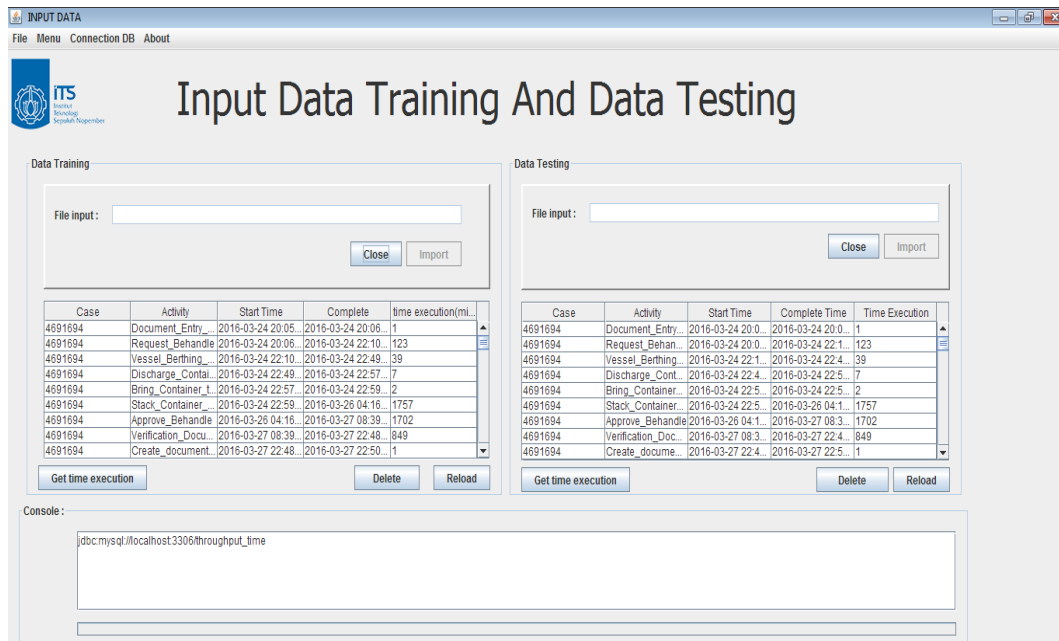
```

Gambar 3.13. Pseudocode for Wrong Decision

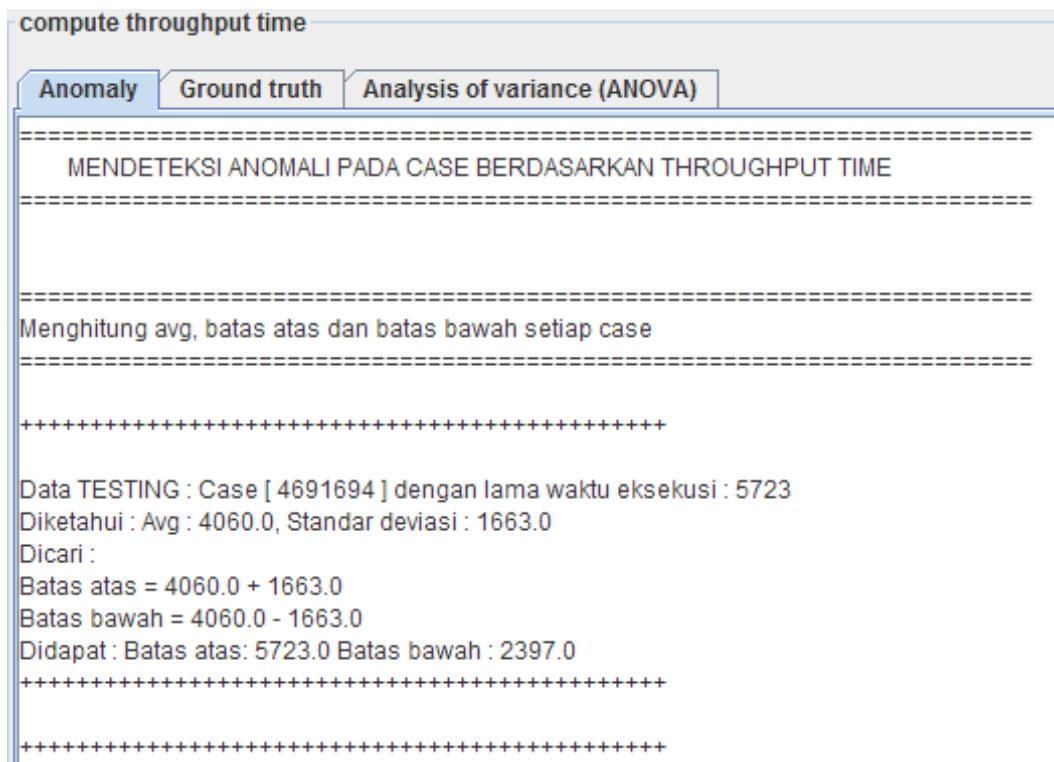
Screenshot Programs untuk mendeteksi *Wrong Throughput Time Min/Max*, *Skip Sequences*, *Wrong Decision* and *Wrong Patterns* ditunjukkan pada gambar 3.14, gambar 3.15, gambar 3.16 dan gambar 3.17.



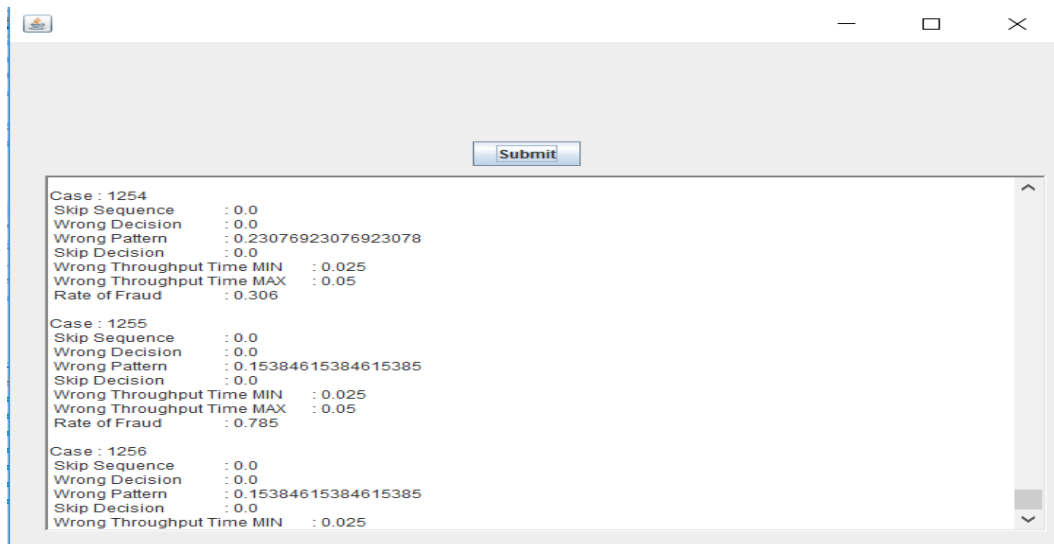
Gambar 3.14. Program Detection Wrong Patterns



Gambar 3.15. Program Detection Input Data Time for Detection Wrong Throughput Time Min/Max



Gambar 3.16. Program Detection Wrong Throughput Time Min/Max



Gambar 3.17. Program Detection Wrong Decisions

3.5 Pendefinisian *Attribute Values*

Setelah dilakukan pendeteksian secara otomatis pada program java yang mana output dari pendeteksian tersebut berupa nilai anomali dari setiap atribut

anomali yang didefinisikan pada program yakni *Wrong Throughput Time Min/Max*, *Skip Sequences*, *Wrong Decision* and *Wrong Patterns*. Setiap atribut anomali memiliki karakteristik atau paramater nilai atribut tersendiri. Berikut pada tabel 3.1 adalah *Attribute Value Skip Sequences (31 Activity)*, tabel 3.2 adalah *Attribute Value Throughput Time Min*, tabel 3.3 adalah *Attribute Value Throughput Time Max*, tabel 3.4 adalah *Attribute Value Wrong Decisions* dan tabel 3.5 adalah *Attribute Value Wrong Patterns*.

Tabel 3.1. Attribute Value Skip Sequences (31 Activity)

| <i>Skip Sequences (31 Activity)</i> | | |
|-------------------------------------|----------------|------------------------|
| <i>Amount of Skip Sequences</i> | <i>Scoring</i> | <i>Attribute Value</i> |
| 1 | 31/31 | 1 |
| 30 | 30/31 | 0.96 |
| ... | ... | ... |
| 0 | 0/31 | 0 |

Tabel 3.2. Attribute Value Throughput Time Min

| <i>Wrong Throughput Time Min (All Activity)</i> | | |
|---|----------------|------------------------|
| <i>Amount of Wrong Throughput Time Min</i> | <i>Scoring</i> | <i>Attribute Value</i> |
| $-SD$ | $-(SD)$ | 1 |
| $-1/2 SD$ | $-1/2.(SD)$ | ... |
| \bar{x} | 0 | 0 |

Tabel 3.3. Attribute Value Throughput Time Max

| <i>Wrong Throughput Time Min (All Activity)</i> | | |
|---|----------------|------------------------|
| <i>Amount of Wrong Throughput Time Min</i> | <i>Scoring</i> | <i>Attribute Value</i> |
| $+SD$ | $+(SD)$ | 1 |
| $+1/2 SD$ | $+1/2.(SD)$ | ... |
| \bar{x} | 0 | 0 |

Tabel 3.4. Attribute Value Wrong Decisions

| <i>Wrong Decisions</i> | | |
|----------------------------------|----------------|------------------------|
| <i>Amount of Wrong Decisions</i> | <i>Scoring</i> | <i>Attribute Value</i> |
| 2 | 2/2 | 1 |
| 1 | 1/2 | 0.50 |
| 0 | 0/2 | 0 |

Tabel 3.5. Attribute Value Wrong Patterns

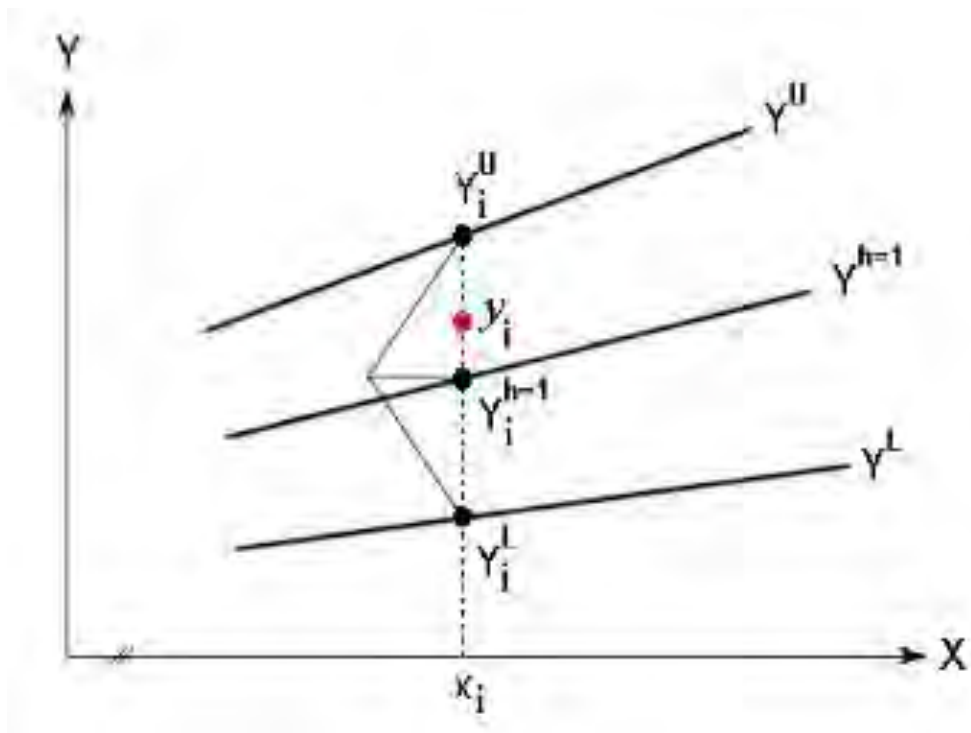
| Wrong Patterns (31 Activity) | | |
|-------------------------------------|----------------|------------------------|
| Amount of Wrong Patterns | Scoring | Attribute Value |
| 30 | 30/30 | 1 |
| ... | ... | ... |
| 1 | 1/30 | 0.03 |
| 0 | 0 | 0 |

3.6 Data Penelitian untuk Proses *Fuzzy Regression*

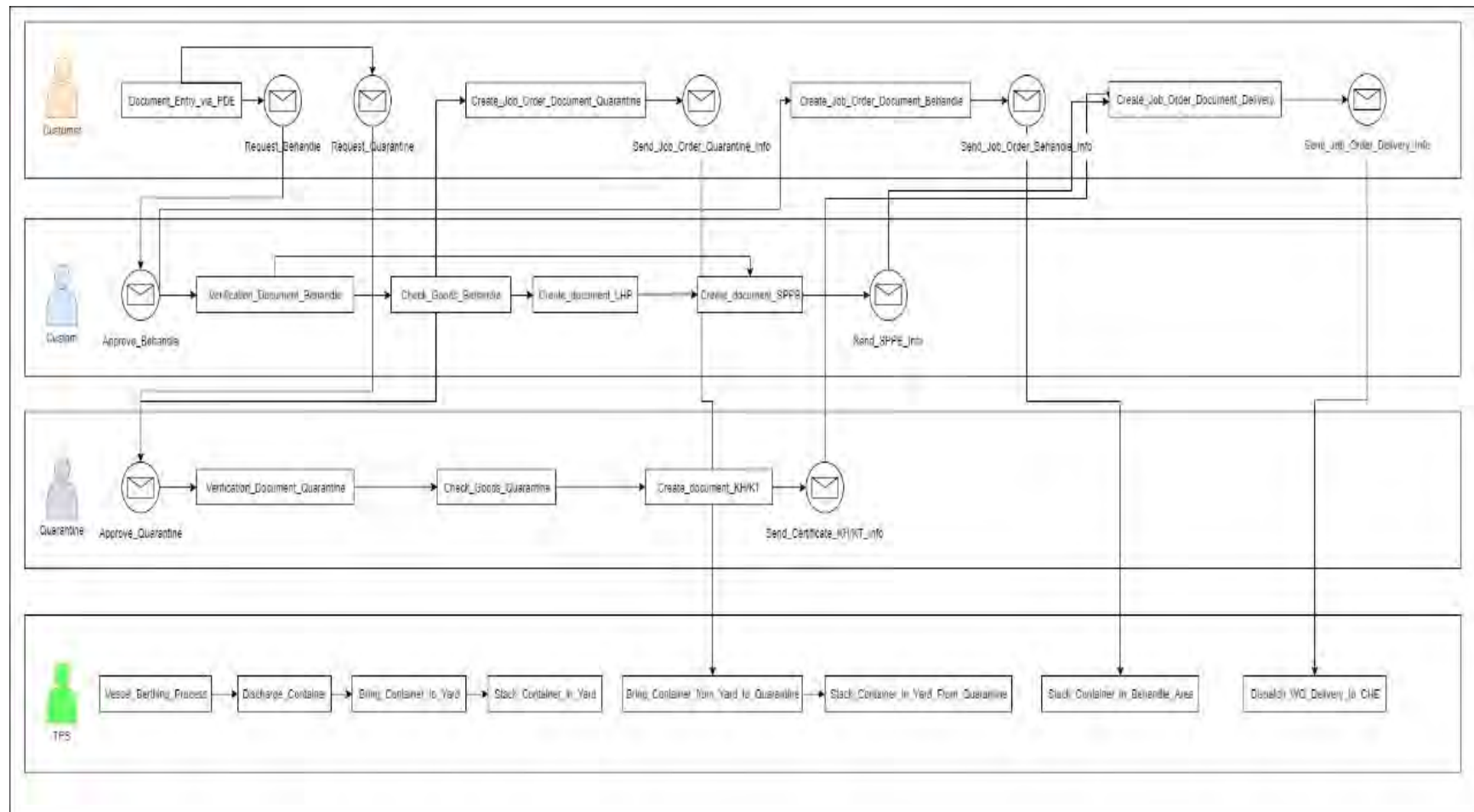
Metode *Fuzzy Regression* digunakan untuk mendapatkan sebuah hubungan linier fuzzy. Penelitian ini bertujuan untuk membandingkan nilai galat yang dihasilkan dari pendugaan model regresi antara metode regresi biasa dan regresi fuzzy. Data yang digunakan dalam penelitian ini adalah data yang berasal dari data log PT.Peti Kemas Surabaya 2015/2016 yang selanjutnya akan dianalisis dan dikonversi menjadi nilai atribut dari kejadian anomali pada setiap trace. Dari hasil penelitian terdapat korelasi yang tinggi antarpeubah, yang mengidentikasi adanya multikolinieritas, sehingga sebelum menggunakan analisis regresi berganda harus melalui tahapan analisis komponen utama. Sementara nilai peubah fuzzy diperoleh dengan menggunakan fungsi keanggotaan segitiga simetris. Diharapkan Hasil penelitian menunjukkan bahwa metode regresi fuzzy relatif lebih baik dibandingkan metode regresi biasa dengan perbedaan galat absolut, MAPE, dan MSE

Ketika Menggunakan Metode Regresi Biasa, hasil kurang akurat, dengan menggunakan regresi fuzzy akan meningkatkan akurasi pendeteksian anomali dari penelitian. Dalam tabel 3.6, data *Attribute of Anomaly* yang berasal dari tabel 3.1, tabel 3.2, tabel 3.3, tabel 3.5 dan *Rate of Anomaly* yang berasal dari PT.Terminal Peti Kemas Surabaya pada Bulan Desember 2015 hingga Maret 2016 yang diperoleh dari Database Event Log PT.Terminal Peti Kemas Surabaya pada tahun 2015-2016 dari beberapa proses event log yakni proses *Discharge*, *Yard*, *Quarantine* dan *Delivery* dengan faktor *Skip Sequence*, *Skip Decision*, *Wrong Throughput Time Min*, *Wrong Throughput Time Max*, dan *Wrong Decision* pada 65000 case (PT.Terminal Peti Kemas Surabaya menerima 15000 *container* / 15000 case per hari) .

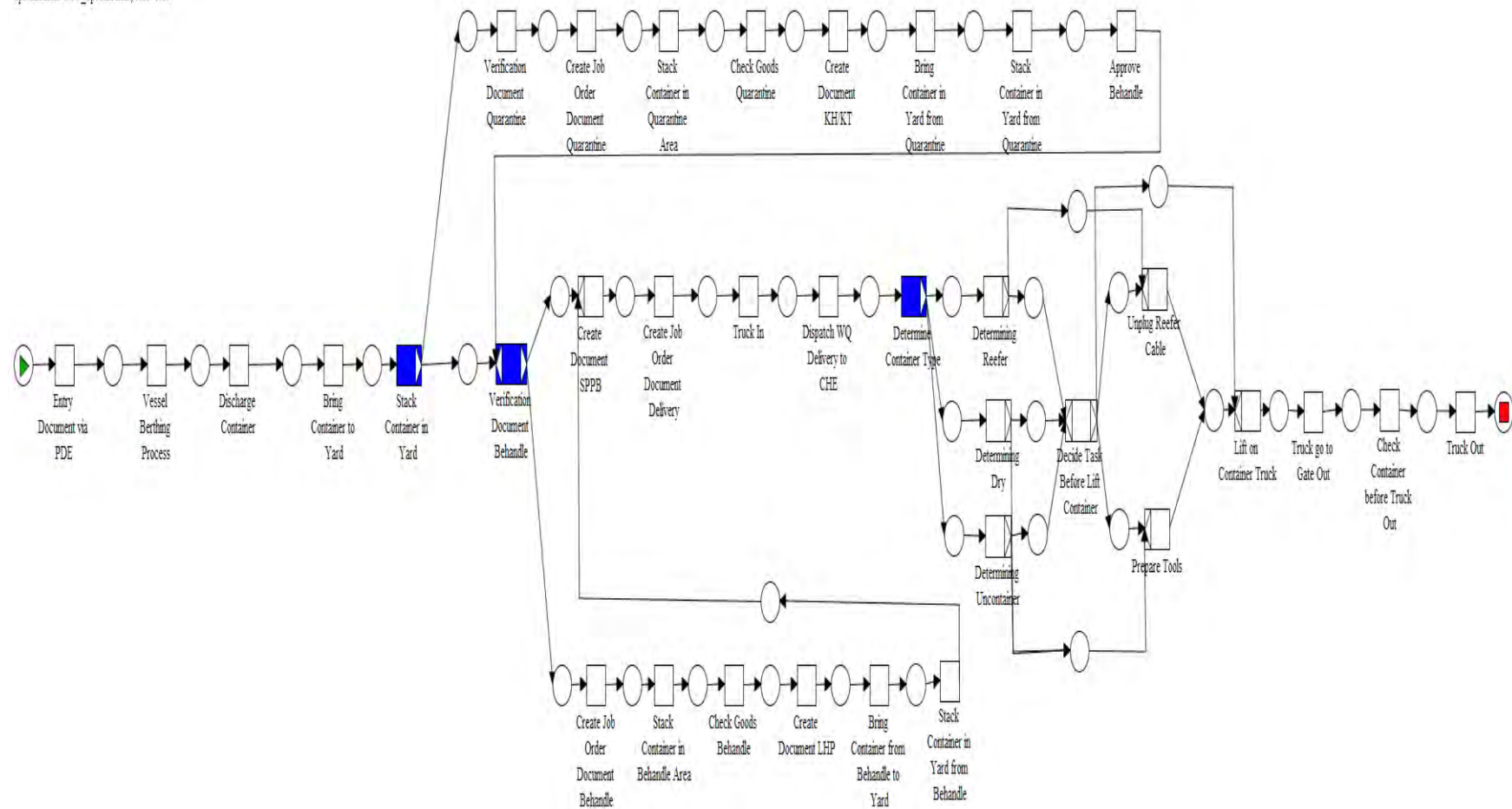
Dengan menggunakan *fuzzy regression*, digabungkan antara metode regresi dan fuzzy sehingga setelah proses regresi dilakukan lalu akan diberi membership function sehingga lebih akurat hasil pendeteksiannya sehingga akan didapatkan model *interval fuzzy regression* seperti gambar 3.18 yang mana merupakan model pembatasan pendeteksian anomali dari indikator *high*, *medium* dan *low* dan Tabel 3.8 adalah tabel yang digunakan untuk proses *Fuzzy Regression* yang mana selanjutnya akan dibuat sebagai bahan untuk perhitungan *Fuzzy Regression*. Gambar 3.19 merupakan gambaran model proses bisnis dari PT. Terminal Peti Kemas (TPS) yang mana terdiri dari proses bisnis dari originator *custom* (Bea Cukai), *customer*, TPS dan *Quarantine* (Petugas Karantina). Dan Gambar 3.20 menjelaskan mengenai pembentukan model proses bisnis orkestrasi dari data event log yang dipunyai oleh TPS yang mana dibentuk melalui tools YAWL (*Yet Another Workflow Language*).



Gambar 3.18. Interval Fuzzy Regression



Gambar 3.19. Gambaran *Business Process* Terminal Peti Kemas (TPS)



Gambar 3.20. Pembentukan Model Proses dengan Yawl Petri Net

Tabel 3.6. Sumber Database 65000 Log (41 kolom label) PT.Terminal Peti Kemas Surabaya

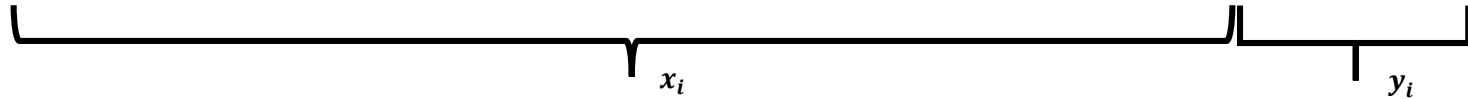
| CONTAINER_KEY | CONTAINER_NO | ISO_CODE | CTR_SIZE | CTR_TYPE | | FIRST_STACK_BEHANDLE_TS | LAST_COMPLETE_BEHANDLE_TS |
|---------------|--------------|----------|----------|----------|-------|-------------------------|---------------------------|
| 3118841 | SGCU2250590 | 22G1 | 20 | DRY | | - | - |
| 3254022 | SGCU5135579 | 45G1 | 40 | DRY | | 8/3/2015 1:51 | 12/3/2015 16:17 |
| 3960409 | FCIU4751761 | 22G1 | 20 | DRY | | 16-09-15 1:08 | 16-09-15 15:59 |
| 4035323 | WHLU2884766 | 2200 | 20 | DRY | | 8/10/2015 0:58 | 8/10/2015 9:56 |
| 4153740 | OOLU7597549 | 42G1 | 40 | DRY | | 14-11-15 1:26 | 17-11-15 11:24 |
| 4153757 | OOLU7767291 | 42G1 | 40 | DRY | | 14-11-15 0:45 | 17-11-15 11:24 |
| 4169037 | AMFU8521170 | 4500 | 40 | DRY | | - | - |
| 4172397 | MRKU4665604 | 4510 | 40 | DRY | | - | - |
| 4172396 | MRKU5071818 | 4510 | 40 | DRY | | - | - |
| 4172389 | CAIU8109561 | 4510 | 40 | DRY | | - | - |
| 4172405 | TGHU8301939 | 4510 | 40 | DRY | | - | - |
| 4172404 | MSKU0463356 | 4510 | 40 | DRY | | - | - |
| 4172402 | MSKU0481940 | 4510 | 40 | DRY | | - | - |
| 4172401 | MSKU8197993 | 4510 | 40 | DRY | | - | - |
| 4172233 | MSKU8310292 | 4510 | 40 | DRY | | - | - |
| 4195501 | CMAU1705181 | 22G1 | 20 | DRY | | - | - |
| 4195496 | CMAU1919542 | 22G1 | 20 | DRY | | - | - |

Tabel 3.7. Sumber Database 600 Log (12 kolom label) PT.Terminal Peti Kemas Surabaya telah terseleksi siap dideteksi

| CASE ID | SENDER | ORIGINATOR | INPUT | AKTIVITAS \ MESSAGE | OUTPUT | RECEIVER | TIME | COST | DETAIL LAMPIRAN |
|---------|----------|------------|--|------------------------------------|-----------|-------------------|------|------|-----------------|
| 3676509 | | CUSTOMER | NPWP, SIUP, API, SRP, TDP, NPIK, IT, INVOICE, PO, SK, BL, COO | Document Entry via PDE | BC 2.0 | | | | "DRY; RED LINE" |
| 3676509 | CUSTOMER | | BC 2.0 | RequestBehandle | BC 2.0 | SKP | | | |
| 3676509 | CUSTOMER | | COO, Health Certificate, Sanitary Certificate, CITES, SPI, BL, Invoice, Packing List, Cargo Manifest, DO | RequestQuarantine | KH-1/KT-1 | Petugas Karantina | | | |
| 3676509 | | TPS | | Vessel Berthing Process | | | | | |
| 3676509 | | TPS | | Discharge Container | | | | | |
| 3676509 | | TPS | | Bring Container to Yard | | | | | |
| 3676509 | | TPS | | Stack Container in Yard | | | | | |
| 3676509 | SKP | | BC 2.0 | Approve Behandle | BC 2.0 | CUSTOMER | | | |
| 3676509 | | SKP | BC 2.0 | Determine Category | BC 2.0 | | | | "DRY; RED LINE" |
| 3676509 | | CUSTOMER | BC 2.0 | Create Job Order Document Behandle | | | | | |
| 3676509 | CUSTOMER | | | SendJobOrderBehandleInfo | | TPS | | | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 3676509 | | CUSTOMER | | Truck Out | | | | | |

Tabel 3.8. Tabel Data Penelitian Rate of Anomaly (From Expert)

| Case | Skip Sequences | Wrong Throughput Time Min | Wrong Throughput Time Max | Wrong Decision | Wrong Pattern | Rate of Anomaly |
|---------------|----------------|---------------------------|---------------------------|----------------|---------------|-----------------|
| 1 | 0 | 0.5 | 0.33 | 0.07 | 0.07 | 0.654 |
| 2 | 0.22 | 0.03 | 0 | 0 | 0.93 | 0.511 |
| 3 | 0 | 0.5 | 0.64 | 0.03 | 0 | 0.900 |
| 4 | 0.03 | 0.5 | 0.45 | 0.02 | 0.02 | 0.729 |
| 5 | 0 | 0.58 | 0 | 0 | 0 | 0.325 |
| 6 | 0.52 | 0.09 | 0 | 0 | 0 | 0.455 |
| 7 | 0.72 | 0.05 | 0 | 0 | 0 | 0.455 |
| 8 | 0 | 0 | 0.37 | 0 | 0 | 0.428 |
| 9 | 0 | 0.03 | 0 | 0 | 0 | 0.353 |
| 10 | 0.21 | 0.06 | 0 | 0 | 0 | 0.418 |
| 11 | 0 | 0.58 | 0 | 0 | 0 | 0.411 |
| 12 | 0 | 0.04 | 0 | 0 | 0 | 0.321 |
| 13 | 0.43 | 0.03 | 0 | 0 | 0 | 0.391 |
| 14 | 0 | 0.58 | 0 | 0 | 0 | 0.328 |
| 15 | 0 | 0.09 | 0 | 0 | 0 | 0.370 |
| 16 | 0.77 | 0.58 | 0 | 0 | 0.57 | 0.375 |
| ... | ... | ... | ... | ... | ... | |
| 639 | 0 | 0 | 0.45 | 0 | 0 | 0.003 |
| 65000 | 0 | 0 | 0.03 | 0 | 0 | 0.003 |
| JUMLAH | 27.92 | 11790.5 | 4427.8 | 499.3 | 155.7 | 3539.4 |



BAB IV

UJI COBA DAN ANALISIS HASIL

4.1 Evaluasi Model dengan *Multiple Linear Regression Model*

Dalam penelitian ini, klasifikasi berbasis regresi terdiri dari satu langkah utama bahwa teknik *Multiple Linear Regression* dievaluasi untuk menentukan tingkat prediksi suatu anomali. Total data kasus adalah 65000 kasus, untuk data pelatihan (data *training*) menggunakan 60% dari 65000 Case of Data dan data pengujian (data *testing*) menggunakan 40% dari 65000 Case of Data. Output regresi sesuai dengan label kelas (“*No Anomaly*”, “*Low Anomaly*”, “*Medium Anomaly*” and “*High Anomaly*”). Dalam thesis ini, Multiple Linear Regression (MLR) memiliki variabel independen x yaitu *Skip Sequences*, *Wrong Throughput Time MIN*, *Wrong Throughput Time MAX*, *Wrong Decision* and *Wrong Patterns Analysis* dan variabel dependen nya adalah y (*Rate of Anomaly predictions*). Variabel x dan y dari *anomaly* selanjutnya digunakan untuk mendapatkan persamaan regresi yang dapat dilihat pada Tabel 4.1 berikut:

Tabel 4.1. Tabel Attribute Value dan Rate of Fraud

| <i>Case</i> | <i>Skip Sequences</i> | <i>Wrong Throughput Time MIN</i> | <i>Wrong Throughput Time MAX</i> | <i>Wrong Decision</i> | <i>Wrong Patterns</i> | <i>Rate of Anomaly</i> |
|-------------|-----------------------|----------------------------------|----------------------------------|-----------------------|-----------------------|------------------------|
| 1 | 0 | 0.5 | 0.33 | 0.07 | 0.07 | 0.654 |
| 2 | 0.22 | 0.03 | 0 | 0 | 0.93 | 0.511 |
| 3 | 0 | 0.5 | 0.64 | 0.03 | 0 | 0.900 |
| 4 | 0.03 | 0.5 | 0.45 | 0.02 | 0.02 | 0.729 |
| ... | ... | ... | ... | ... | ... | ... |
| 65000 | 0 | 0 | 0.03 | 0 | 0 | 0.003 |

Untuk mendeteksi mana yang bukan merupakan anomali, anomali rendah, anomali menengah dan kelas anomali tinggi, ahli menentukan *threshold* dimana ahli menentukan di atas 0,6 sebagai anomali dan kemudian bahwa dengan *threshold* ditemukan 680 kasus yang mengandung anomali (yang memiliki Tingkat Anomali (ROA) di atas 0,6). Pengkategorian ROA (Rate of Anomaly) seperti tabel 4.2 berikut:

Tabel 4.2. Category of Anomaly in Multiple Linear Regression

| <i>Class Label</i> | <i>Range of Rate of Anomaly</i> | <i>Number of Cases</i> |
|--------------------|---------------------------------|------------------------|
| High Anomaly | $0.8 \leq x \leq 1$ | 80 |
| Medium Anomaly | $0.7 \leq x < 0.8$ | 280 |
| Low Anomaly | $0.6 \leq x < 0.7$ | 320 |
| No Anomaly | $0 \leq x < 0.6$ | 64320 |

Data telah diberi label dan ada 4 kelas anomali tinggi, anomali medium, anomali rendah dan tidak ada anomali. Kemudian tentukan data pelatihan dan pengujian data dengan mengambil data dari masing - masing kelas seperti yang ditunjukkan pada Tabel 4.3.

Tabel 4.3. Number of Cases of Training and Testing Multiple Linear Regression

| <i>Class Label</i> | <i>Number of Cases for Training</i> | <i>Number of Cases for Testing</i> |
|--------------------|-------------------------------------|------------------------------------|
| High Anomaly | 48 | 32 |
| Medium Anomaly | 168 | 112 |
| Low Anomaly | 192 | 128 |
| No Anomaly | 38592 | 25728 |

Data pelatihan (*training*) digunakan untuk membangun model *Multiple Linear Regression* dengan x terdiri dari 5 nilai atribut dan y adalah rating anomali. Kemudian dengan alat *Minitab* menemukan model *Multiple Linear Regression*. Data training *ROAHIGH*, *ROAMED*, *ROALOW*, dan *ROANO* didapatkan dari variabel input seperti pada Tabel 4.4, Tabel 4.5, Tabel 4.6 dan Tabel 4.7 berikut:

Tabel 4.4. Data Training ROAHIGH

| Case | SS | WD | WP | TMIN | TMAX | ROA |
|-------------|-----------|-----------|-----------|-------------|-------------|------------|
| 1 | 0 | 0.5 | 0.741935 | 0.03125 | 0 | 1 |
| 2 | 0 | 0.5 | 0.741935 | 0.03125 | 0 | 1 |
| 3 | 0 | 0.5 | 0.645161 | 0.03125 | 0 | 0.900651 |
| 4 | 0 | 0.5 | 0.645161 | 0.03125 | 0 | 0.900651 |
| 5 | 0 | 0.5 | 0.645161 | 0.03125 | 0 | 0.900651 |
| 6 | 0 | 0.5 | 0.645161 | 0.03125 | 0 | 0.900651 |
| 7 | 0 | 0.5 | 0.645161 | 0.03125 | 0 | 0.900651 |

| Case | SS | WD | WP | TMIN | TMAX | ROA |
|------|----------|-----|----------|----------|----------|----------|
| 8 | 0 | 0.5 | 0.645161 | 0.03125 | 0 | 0.900651 |
| 9 | 0 | 0.5 | 0.645161 | 0.03125 | 0 | 0.900651 |
| 10 | 0 | 0.5 | 0.645161 | 0.03125 | 0 | 0.900651 |
| ... | ... | ... | ... | ... | ... | ... |
| 39 | 0.030303 | 0.5 | 0.545455 | 0.029412 | 0.029412 | 0.823224 |
| 40 | 0.030303 | 0.5 | 0.545455 | 0.029412 | 0.029412 | 0.823224 |
| 41 | 0.030303 | 0.5 | 0.545455 | 0.029412 | 0.029412 | 0.823224 |
| 42 | 0.030303 | 0.5 | 0.545455 | 0.029412 | 0.029412 | 0.823224 |
| 43 | 0.030303 | 0.5 | 0.545455 | 0.029412 | 0.029412 | 0.823224 |
| 44 | 0.030303 | 0.5 | 0.545455 | 0.029412 | 0.029412 | 0.823224 |
| 45 | 0.030303 | 0.5 | 0.545455 | 0.029412 | 0.029412 | 0.823224 |
| 46 | 0 | 0.5 | 0.571429 | 0 | 0 | 0.806623 |
| 47 | 0 | 0.5 | 0.548387 | 0.03125 | 0 | 0.801301 |
| 48 | 0 | 0.5 | 0.548387 | 0.03125 | 0 | 0.801301 |

Tabel 4.5. Data Training ROAMED

| Case | SS | WD | WP | TMIN | TMAX | ROA |
|------|----------|-----|----------|----------|----------|----------|
| 1 | 0.030303 | 0.5 | 0.515152 | 0.029412 | 0.029412 | 0.792114 |
| 2 | 0.030303 | 0.5 | 0.515152 | 0.029412 | 0.029412 | 0.792114 |
| 3 | 0.030303 | 0.5 | 0.515152 | 0.029412 | 0.029412 | 0.792114 |
| 4 | 0.030303 | 0.5 | 0.515152 | 0.029412 | 0.029412 | 0.792114 |
| 5 | 0.030303 | 0.5 | 0.515152 | 0.029412 | 0.029412 | 0.792114 |
| 6 | 0.030303 | 0.5 | 0.515152 | 0.029412 | 0.029412 | 0.792114 |
| 7 | 0.030303 | 0.5 | 0.515152 | 0.029412 | 0.029412 | 0.792114 |
| 8 | 0.030303 | 0.5 | 0.515152 | 0.029412 | 0.029412 | 0.792114 |
| ... | ... | ... | ... | ... | ... | ... |
| 156 | 0.030303 | 0.5 | 0.454545 | 0.029412 | 0.029412 | 0.729895 |
| 157 | 0.030303 | 0.5 | 0.454545 | 0.029412 | 0.029412 | 0.729895 |
| 158 | 0.030303 | 0.5 | 0.454545 | 0.029412 | 0.029412 | 0.729895 |
| 159 | 0 | 0.5 | 0.451613 | 0.03125 | 0 | 0.711561 |
| 160 | 0.025641 | 0.5 | 0.410256 | 0.05 | 0.05 | 0.710919 |
| 161 | 0 | 0.5 | 0.47619 | 0 | 0 | 0.708851 |
| 162 | 0 | 0.5 | 0.47619 | 0 | 0 | 0.708851 |
| 163 | 0 | 0.5 | 0.47619 | 0 | 0 | 0.708851 |
| 164 | 0 | 0.5 | 0.451613 | 0.03125 | 0 | 0.701952 |
| 165 | 0 | 0.5 | 0.451613 | 0.03125 | 0 | 0.701952 |
| 166 | 0 | 0.5 | 0.451613 | 0.03125 | 0 | 0.701952 |
| 167 | 0 | 0.5 | 0.451613 | 0.03125 | 0 | 0.701952 |
| 168 | 0 | 0.5 | 0.451613 | 0.03125 | 0 | 0.701952 |

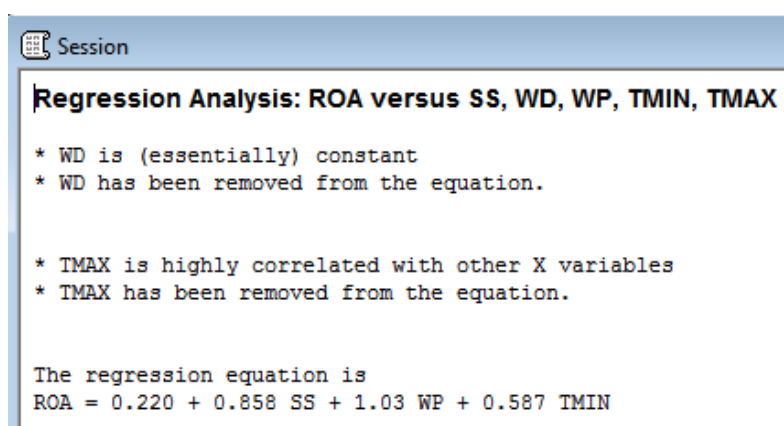
Tabel 4.6. Data Training ROALOW

| Case | SS | WD | WP | TMIN | TMAX | ROA |
|------|----------|-----|----------|----------|----------|----------|
| 1 | 0.025641 | 0.5 | 0.384615 | 0.05 | 0.05 | 0.684596 |
| 2 | 0.025641 | 0.5 | 0.384615 | 0.05 | 0.05 | 0.684596 |
| 3 | 0 | 0.5 | 0.45 | 0 | 0 | 0.681963 |
| 4 | 0 | 0.5 | 0.45 | 0 | 0 | 0.681963 |
| 5 | 0 | 0.5 | 0.45 | 0 | 0 | 0.681963 |
| 6 | 0 | 0.5 | 0.45 | 0 | 0 | 0.681963 |
| 7 | 0 | 0.5 | 0.45 | 0 | 0 | 0.681963 |
| ... | ... | ... | ... | ... | ... | ... |
| 185 | 0.030303 | 0.5 | 0.363636 | 0.029412 | 0.029412 | 0.636567 |
| 186 | 0.030303 | 0.5 | 0.363636 | 0.029412 | 0.029412 | 0.636567 |
| 187 | 0.030303 | 0.5 | 0.363636 | 0.029412 | 0.029412 | 0.636567 |
| 188 | 0.030303 | 0.5 | 0.363636 | 0.029412 | 0.029412 | 0.636567 |
| 189 | 0.030303 | 0.5 | 0.363636 | 0.029412 | 0.029412 | 0.636567 |
| 190 | 0.030303 | 0.5 | 0.363636 | 0.029412 | 0.029412 | 0.636567 |
| 191 | 0.030303 | 0.5 | 0.363636 | 0.029412 | 0.029412 | 0.636567 |
| 192 | 0.030303 | 0.5 | 0.363636 | 0.029412 | 0.029412 | 0.636567 |

Tabel 4.7. Data Training ROANo

| Case | SS | WD | WP | TMIN | TMAX | ROA |
|-------|----------|-----|----------|----------|----------|-------------|
| 1 | 0.030303 | 0.5 | 0.363636 | 0.029412 | 0.029412 | 0.636567084 |
| 2 | 0.030303 | 0 | 0 | 0 | 0 | 0.004444205 |
| 21 | 0 | 0.5 | 0 | 0 | 0 | 0.219988173 |
| 39 | 0 | 0 | 0.363636 | 0 | 0 | 0.373313263 |
| 62 | 0 | 0 | 0 | 0.029412 | 0 | 0.017253974 |
| 84 | 0 | 0 | 0 | 0 | 0.029412 | 0.021567468 |
| 103 | 0 | 0 | 0 | 0.029412 | 0.029412 | 0.038821443 |
| 118 | 0 | 0 | 0.363636 | 0.029412 | 0 | 0.390567238 |
| 137 | 0.030303 | 0.5 | 0 | 0 | 0 | 0.224432378 |
| ... | ... | ... | ... | ... | ... | ... |
| 64312 | 0 | 0 | 0 | 0 | 0 | 0 |
| 64313 | 0 | 0 | 0 | 0 | 0 | 0 |
| 64314 | 0 | 0 | 0 | 0 | 0 | 0 |
| 64315 | 0 | 0 | 0 | 0 | 0 | 0 |
| 64316 | 0 | 0 | 0 | 0 | 0 | 0 |
| 64317 | 0 | 0 | 0 | 0 | 0 | 0 |
| 64318 | 0 | 0 | 0 | 0 | 0 | 0 |
| 64319 | 0 | 0 | 0 | 0 | 0 | 0 |
| 64320 | 0 | 0 | 0 | 0 | 0 | 0 |

Kemudian untuk mendapatkan persamaan *ROAHIGH*, *ROAMED*, *ROALOW* dan *ROANO* digunakan tool Minitab yang mana dengan menggunakan data input masing-masing pembagian data training seperti pada Tabel 4.4, Tabel 4.5, Tabel 4.6 dan Tabel 4.7 dan menghasilkan Persamaan (8), (9), (10) dan (11) dari hasil analisis pada tool minitab seperti pada Gambar 4.1, Gambar 4.2, Gambar 4.3 dan Gambar 4.4 berikut:



```

Session

Regression Analysis: ROA versus SS, WD, WP, TMIN, TMAX

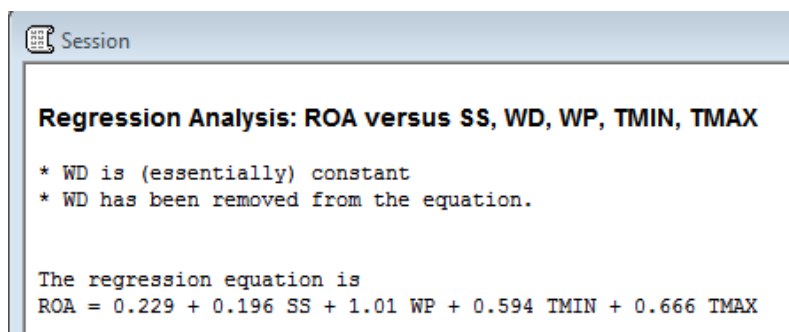
* WD is (essentially) constant
* WD has been removed from the equation.

* TMAX is highly correlated with other X variables
* TMAX has been removed from the equation.

The regression equation is
ROA = 0.220 + 0.858 SS + 1.03 WP + 0.587 TMIN

```

Gambar 4.1. Result ROAHigh Regression Analysis with Minitab Tool



```

Session

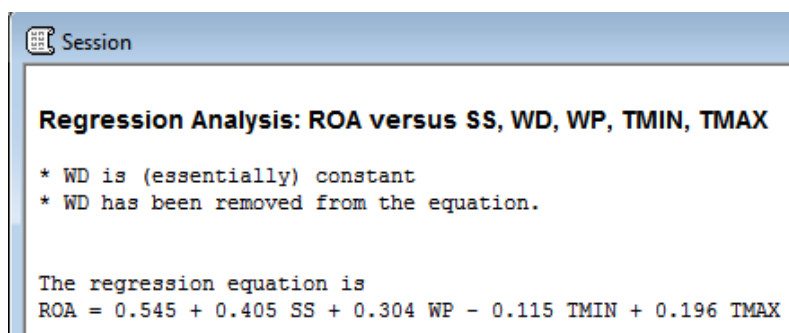
Regression Analysis: ROA versus SS, WD, WP, TMIN, TMAX

* WD is (essentially) constant
* WD has been removed from the equation.

The regression equation is
ROA = 0.229 + 0.196 SS + 1.01 WP + 0.594 TMIN + 0.666 TMAX

```

Gambar 4.2. Result ROAMED Regression Analysis with Minitab Tool



```

Session

Regression Analysis: ROA versus SS, WD, WP, TMIN, TMAX

* WD is (essentially) constant
* WD has been removed from the equation.

The regression equation is
ROA = 0.545 + 0.405 SS + 0.304 WP - 0.115 TMIN + 0.196 TMAX

```

Gambar 4.3. Result ROALOW Regression Analysis with Minitab Tool

| Session | |
|--|--|
| * NOTE * All values in column are identical.* NOTE * All values in column are identical. | |
| Regression Analysis: ROA versus SS, WD, WP, TMIN, TMAX | |
| The regression equation is | |
| ROA = 0.000000 + 0.147 SS + 0.440 WD + 1.03 WP + 0.587 TMIN + 0.733 TMAX | |

Gambar 4.4. Result ROANO Regression Analysis with Minitab Tool

Hasil *Multiple Linear Regression* seperti Gambar 4.1, Gambar 4.2, Gambar 4.3, Gambar 4.4 dan Persamaan (8), (9), (10) dan (11). Model *Multiple Linear Regression* dibangun untuk menemukan prediksi tingkat anomali dari semua nilai atribut anomali yang nilainya *continuous* (0.1, 0.2, ..., Xn).

$$ROAHIGH = 0.220 + 0.858 SS + 1.03 WP + 0.587 TMIN \quad (8)$$

$$ROAMED = 0.229 + 0.196 SS + 1.01 WP + 0.594 TMIN + 0.666 TMAX \quad (9)$$

$$ROALOW = 0.545 + 0.405 SS + 0.304 WP - 0.115 TMIN + 0.196 TMAX \quad (10)$$

$$ROANO = 0 + 0.147 SS + 0.440 WD + 1.03 WP + 0.587 TMIN + 0.733 TMAX \quad (11)$$

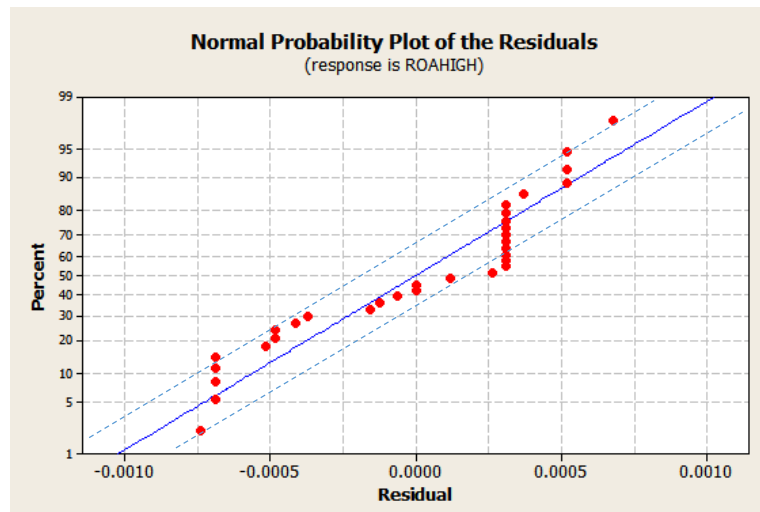
Setelah mendapatkan model persamaan dari data *training* kemudian sekarang men-*testing data* dengan data testing per masing-masing kriteria. Berikut adalah contoh proses mentesting data dari data *high anomaly* yang diketahui dari *threshold high* seperti pada Tabel 4.8 berikut:

Tabel 4.8. Testing Data High Anomaly

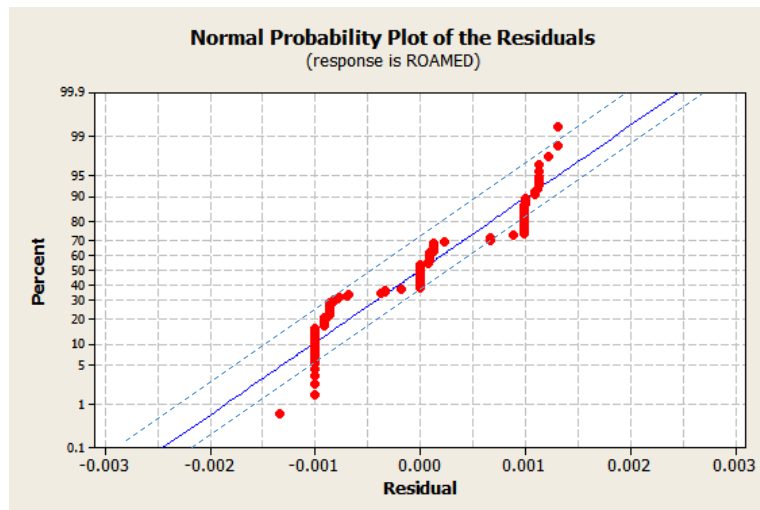
| Case | SS | WD | WP | TMIN | TMAX | ROA | Hasil Testing |
|------|-------|-----|------|----------|----------|----------|---------------|
| 1 | 0 | 0.5 | 0.61 | 0.03125 | 0 | 0.869634 | true |
| 2 | 0.030 | 0.5 | 0.57 | 0.029412 | 0.029412 | 0.856295 | true |
| 3 | 0.030 | 0.5 | 0.57 | 0.029412 | 0.029412 | 0.856295 | true |
| 4 | 0.030 | 0.5 | 0.57 | 0.029412 | 0.029412 | 0.856295 | true |
| 5 | 0.030 | 0.5 | 0.57 | 0.029412 | 0.029412 | 0.856295 | true |
| 6 | 0.030 | 0.5 | 0.57 | 0.029412 | 0.029412 | 0.856295 | true |
| 7 | 0.030 | 0.5 | 0.57 | 0.029412 | 0.029412 | 0.856295 | true |
| 8 | 0.030 | 0.5 | 0.57 | 0.029412 | 0.029412 | 0.856295 | true |
| 9 | 0.030 | 0.5 | 0.57 | 0.029412 | 0.029412 | 0.856295 | true |
| 10 | 0.030 | 0.5 | 0.57 | 0.029412 | 0.029412 | 0.856295 | true |
| 11 | 0 | 0.5 | 0.6 | 0 | 0 | 0.838 | true |
| 12 | 0 | 0.5 | 0.58 | 0.03125 | 0 | 0.836408 | true |
| 13 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 14 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |

| Case | SS | WD | WP | TMIN | TMAX | ROA | HasilTesting |
|------|-------|-----|------|----------|----------|----------|--------------|
| 15 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 16 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 17 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 18 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 19 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 20 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 21 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 22 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 23 | 0.030 | 0.5 | 0.54 | 0.029412 | 0.029412 | 0.825083 | true |
| 24 | 0 | 0.5 | 0.57 | 0 | 0 | 0.808571 | true |
| 25 | 0 | 0.5 | 0.54 | 0.03125 | 0 | 0.803182 | true |
| 26 | 0 | 0 | 0 | 0 | 0 | 0.62 | wrong |
| 27 | 0 | 0 | 0 | 0 | 0 | 0.62 | wrong |
| 28 | 0 | 0 | 0 | 0 | 0 | 0.72 | wrong |
| 29 | 0 | 0 | 0 | 0 | 0 | 0.72 | wrong |
| 30 | 0 | 0 | 0 | 0 | 0 | 0.72 | wrong |
| 31 | 0 | 0 | 0 | 0 | 0 | 0.72 | wrong |
| 32 | 0 | 0 | 0 | 0 | 0 | 0.72 | wrong |

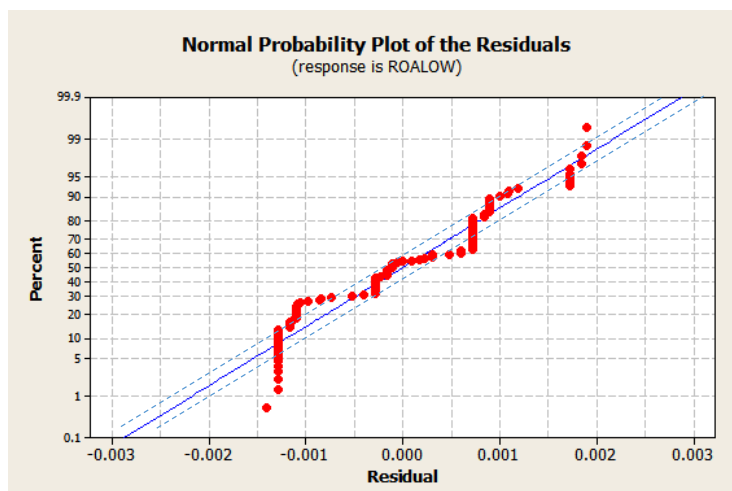
Data mendeteksi benar *High Anomaly* ketika antara data *predicted rate of anomaly* yang diperoleh dari model itu sama dengan data *ground truth* yang diperoleh dari *expert*. Dalam data testing *high anomaly*, setelah diterapkan persamaan (8), *Case* 1-25 memenuhi range *high anomaly* yang mana diantara range $0.8 \leq x \leq 1$, sedangkan *case* 26-32 tidak memenuhi range *high anomaly* (angka dibawah *threshold*). Maka dari itu terdapat 7 salah pendeteksian yang mana 5 dari medium dan 2 dari low. Dan alur pendeteksian ini diterapkan pada *Rate of Anomaly Medium*, *Rate of Anomaly Low* dan *Rate of Anomaly No* juga. Dan hasilnya seperti pada Tabel 4.9. Pada Gambar 4.5, Gambar 4.6, Gambar 4.7 dan Gambar 4.8 merupakan plot pendeteksian *High Anomaly*, *Medium Anomaly*, *Low Anomaly* dan *No Anomaly* yang mana didapatkan dari model *Multiple Linear Regression*.



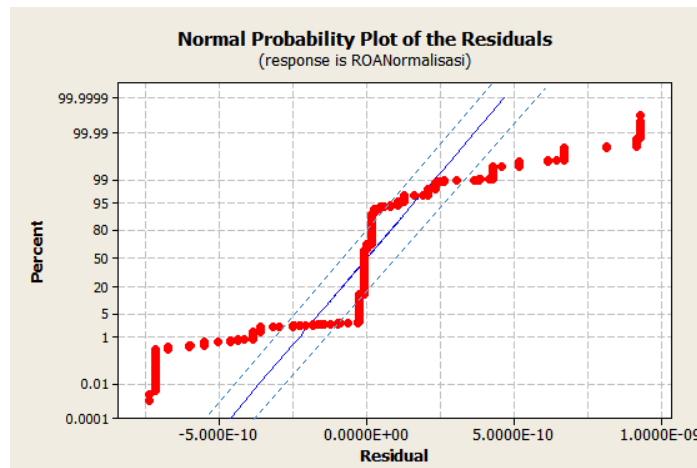
Gambar 4.5. MLR Model of ROA (Rate of Anomaly) High Anomaly



Gambar 4.6. MLR Model of ROA (Rate of Anomaly) Medium Anomaly



Gambar 4.7. MLR Model of ROA (Rate of Anomaly) Low Anomaly



Gambar 4.8. MLR Model of ROA (Rate of Anomaly) No Anomaly

Total data kasus adalah 65000 kasus, untuk data pelatihan menggunakan 60% dari 65000 Case of Data dan data pengujian menggunakan 40% dari 65000 Case of Data. Persamaan itu (4) dibangun dari data pelatihan, setelah itu, uji data dengan persamaan (4) untuk mengetahui tingkat anomali, jika tingkat anomali pengujian lebih besar sama dengan tingkat anomali dalam latihan, kasusnya adalah Mendeteksi anomali kanan (Prediksi Kanan) dan jika tingkat anomali pengujian kurang sama maka tingkat anomali dalam latihan, kasusnya salah atau salah mendeteksi anomali (Prediksi Palsu). Kemudian hitunglah prediksi dengan mencocokkan label dimana label sebagai ground truth untuk mendapatkan nilai deteksi yang dimasukkan ke tabel 4.9 (*Confusion Matrix for Multiple Linear Regression*) dan nilai sensitivitas dan spesifisitas yang diperoleh pada masing-masing label.

Tabel 4.9. Confusion Matrix for Multiple Linear Regression

| EXPERT SYSTEM | HIGH | MEDIUM | LOW | NO ANOMALY |
|------------------|------|--------|-----|---------------|
| HIGH | 25 | 5 | 2 | 0 |
| MEDIUM | 5 | 90 | 10 | 7 |
| LOW | 0 | 15 | 95 | 18 |
| NO ANOMALY | 28 | 200 | 500 | 25000 |
| SENSITIVITY | 43% | 29% | 15% | 99% |
| SPECIFICITY | 97% | 97% | 98% | 21% |

4.2 Evaluasi Model dengan *Fuzzy Regression Model*

Analisis regresi lebih sesuai untuk sistem yang menghasilkan output kontinyu. Dalam penelitian ini, akan digunakan output diskrit karena data dari ahlinya bukan keluaran kontinyu melainkan keluaran diskrit (rendah, sedang dan tinggi) untuk klasifikasi suatu anomali. Pada penelitian sebelumnya, model regresi fuzzy adalah yang terbaik untuk mengklasifikasikan klasifikasi dengan menggunakan beberapa dataset. Regresi fuzzy diperkenalkan sebagai alternatif untuk mengatasi masalah klasifikasi. Hasil percobaan menunjukkan bahwa model regresi fuzzy adalah yang terbaik untuk mengklasifikasikan suatu masalah yang terbukti dengan hasil yang lebih baik dengan menggunakan pengujian data. Hasil ini menunjukkan bahwa klasifikasi regresi fuzzy menghasilkan model overfit. Model Regresi Fuzzy dibangun berdasarkan sistem inferensi fuzzy tipe Mamdani, dimana aturan mengadopsi bentuk berikut seperti pada persamaan (12):

$$\begin{aligned} & \text{IF } X_1 \text{ is } A_{1j} \text{ AND } X_2 \text{ is } A_{2j} \text{ AND } \dots \text{ AND } X_n \text{ is } A_{nj} \\ & \text{THEN } y = c_0 + c_1x_1 + C_2x_2 + \dots + c_nx_n \end{aligned} \quad (12)$$

Dimana A_{1j} adalah himpunan fuzzy yang sesuai dengan j_{th} istilah linguistik dari i_{th} variable, n adalah jumlah variabel input (yaitu, 6 input dalam penelitian ini), y adalah nilai prediksi tingkat anomali dan menunjukkan parameter konsekuen yang nilainya berasal dari proses pelatihan. Langkah pertama adalah membangun keanggotaan 5 atribut yaitu *Skip Sequences* seperti pada tabel 4.10, *Wrong Throughput Time MIN* seperti pada tabel 4.11, *Wrong Throughput Time MAX* seperti pada tabel 4.12, *Wrong Decision* seperti pada tabel 4.13, *Wrong Patterns* seperti pada tabel 4.14 dan tabel *Rate of Anomaly* seperti pada tabel 4.15.

Tabel 4.10. Table Fuzzy Skip Sequences

| <i>SKIP SEQUENCES</i> | | |
|-----------------------|----------------------|----------------------|
| <i>SCORING</i> | <i>FUZZIFICATION</i> | <i>NORMALIZATION</i> |
| > 0.03 | HIGH | $x \geq 1$ |
| $0.02 - 0.03$ | MEDIUM | $0.1 \leq x < 1$ |
| $0 < x < 0.02$ | LOW | $0 < x < 0.1$ |
| ≤ 0 | NO | 0 |

Tabel 4.11. Table Fuzzy Wrong Throughput Time Min

| WRONG THROUGHPUT TIME MIN | | |
|-------------------------------------|----------------------|----------------------|
| SCORING | FUZZIFICATION | NORMALIZATION |
| – (SD) | HIGH | $x \geq 1$ |
| $-\frac{1}{2} \cdot (SD)$ | MEDIUM | $0.1 \leq x < 1$ |
| $0 < x < (-\frac{1}{2} \cdot (SD))$ | LOW | $0 < x < 0.1$ |
| ≤ 0 | No | 0 |

Tabel 4.12. Table Fuzzy Wrong Throughput Time Max

| WRONG THROUGHPUT TIME MAX | | |
|-------------------------------------|----------------------|----------------------|
| SCORING | FUZZIFICATION | NORMALIZATION |
| + (SD) | HIGH | $x \geq 1$ |
| $+\frac{1}{2} \cdot (SD)$ | MEDIUM | $0.1 \leq x < 1$ |
| $0 < x < (+\frac{1}{2} \cdot (SD))$ | LOW | $0 < x < 0.1$ |
| ≤ 0 | No | 0 |

Tabel 4.13. Table Fuzzy Wrong Decisions

| WRONG DECISIONS | | |
|------------------------|----------------------|----------------------|
| SCORING | FUZZIFICATION | NORMALIZATION |
| >0.5 | HIGH | $x \geq 1$ |
| $0.4 - 0.5$ | MEDIUM | $0.1 \leq x < 1$ |
| $0 < x < 0.4$ | LOW | $0 < x < 0.1$ |
| ≤ 0 | No | 0 |

Tabel 4.14. Table Fuzzy Wrong Pattern

| WRONG PATTERNS | | |
|-----------------------|----------------------|----------------------|
| SCORING | FUZZIFICATION | NORMALIZATION |
| >0.35 | HIGH | $x \geq 1$ |
| $0.23 - 0.35$ | MEDIUM | $0.1 \leq x < 1$ |
| $0 < x < 0.23$ | LOW | $0 < x < 0.1$ |
| ≤ 0 | No | 0 |

Tabel 4.15. Table Fuzzy Rate of Anomaly

| RATE OF ANOMALY | | |
|------------------------|----------------------|----------------------|
| SCORING | FUZZIFICATION | NORMALIZATION |
| >0.79 | HIGH | $x \geq 1$ |
| $0.20 - 0.79$ | MEDIUM | $0.1 \leq x < 1$ |
| $0 < x < 0.20$ | LOW | $0 < x < 0.1$ |
| ≤ 0 | No | 0 |

Langkah selanjutnya adalah membagi nilai atribut menjadi anomali kriteria rendah, anomali kriteria medium dan anomali kriteria tinggi untuk mengetahui tingkat fuzzy masing-masing kriteria seperti pada Tabel 4.16, Tabel 4.17, Tabel 4.18 dan Tabel 4.19.

Tabel 4.16. Low Anomaly of 5 Attributes

| <i>Low Anomaly</i> | | | | | |
|----------------------|-----------------------|----------------------|----------------------------------|----------------------------------|------------------------|
| <i>Skip Sequence</i> | <i>Wrong Decision</i> | <i>Wrong Pattern</i> | <i>Wrong Throughput Time Min</i> | <i>Wrong Throughput Time Max</i> | <i>Rate of Anomaly</i> |
| 0 | 0 | 0.27 | 0.09 | 0.12 | 0.19 |
| 0 | 0 | 0.27 | 0.09 | 0 | 0.19 |
| ... | ... | ... | ... | ... | ... |
| 0 | 0 | 0 | 0 | 0 | 0.001 |

Tabel 4.17. Medium Anomaly of 5 Attributes

| <i>Medium Anomaly</i> | | | | | |
|-----------------------|-----------------------|----------------------|----------------------------------|----------------------------------|------------------------|
| <i>Skip Sequence</i> | <i>Wrong Decision</i> | <i>Wrong Pattern</i> | <i>Wrong Throughput Time Min</i> | <i>Wrong Throughput Time Max</i> | <i>Rate of Anomaly</i> |
| 0.10 | 0.37 | 0.55 | 0.20 | 0.49 | 0.79 |
| 0 | 0.37 | 0.48 | 0.13 | 0.49 | 0.79 |
| ... | ... | ... | ... | ... | ... |
| 0 | 0 | 0 | 0 | 0 | 0.28 |

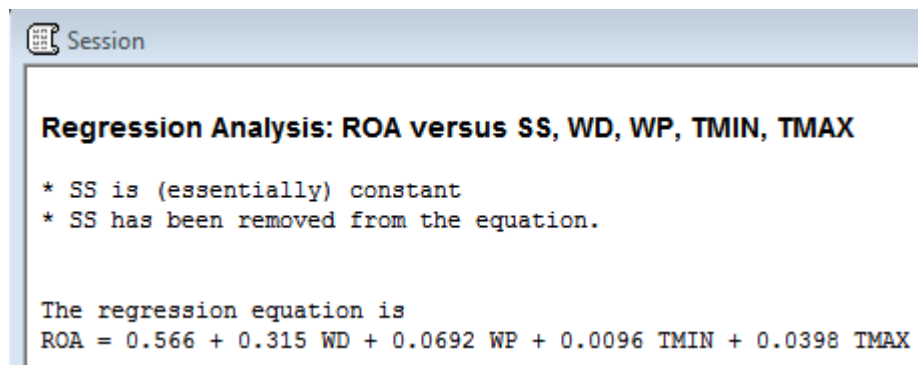
Tabel 4.18. High Anomaly of 5 Attributes

| <i>High Anomaly</i> | | | | | |
|----------------------|-----------------------|----------------------|----------------------------------|----------------------------------|------------------------|
| <i>Skip Sequence</i> | <i>Wrong Decision</i> | <i>Wrong Pattern</i> | <i>Wrong Throughput Time Min</i> | <i>Wrong Throughput Time Max</i> | <i>Rate of Anomaly</i> |
| 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 0.87 | 1 | 1 | 1 |
| ... | ... | ... | ... | ... | ... |
| 0 | 0 | 0 | 0 | 0 | 0.79 |

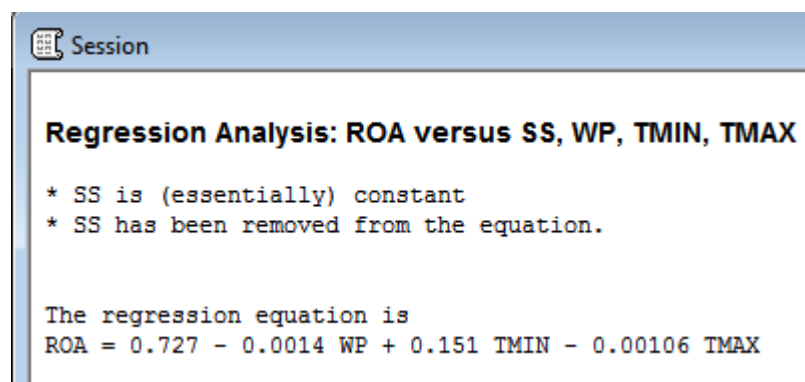
Tabel 4.19. No Anomaly of 5 Attributes

| <i>High Anomaly</i> | | | | | |
|----------------------|-----------------------|----------------------|----------------------------------|----------------------------------|------------------------|
| <i>Skip Sequence</i> | <i>Wrong Decision</i> | <i>Wrong Pattern</i> | <i>Wrong Throughput Time Min</i> | <i>Wrong Throughput Time Max</i> | <i>Rate of Anomaly</i> |
| 0 | 0 | 0 | 0 | 0 | 0.594908187 |
| 0 | 0 | 0 | 0 | 0 | 0.594908187 |
| ... | ... | ... | ... | ... | ... |
| 0 | 0 | 0 | 0 | 0 | 0 |

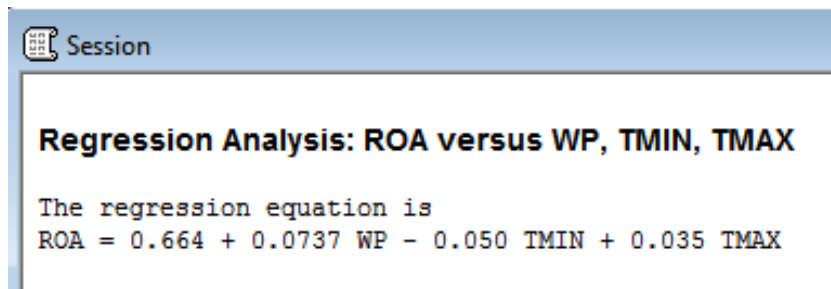
Kemudian langkah selanjutnya adalah menggunakan algoritma *Multiple Linear Regression* dengan hasil data yang sudah di-fuzzy-kan untuk mengetahui persamaan *fuzzy no anomaly*, *fuzzy anomali rendah*, *fuzzy anomali medium* dan *fuzzy anomali tinggi* untuk data *training*. Persamaan (13), (14), (15) dan (16) adalah hasil dari persamaan *fuzzy no anomaly*, *fuzzy anomali rendah*, *fuzzy anomali medium* dan *fuzzy anomali tinggi*. Untuk mendapatkan persamaan *ROAHIGH*, *ROAMED*, *ROALOW* dan *ROANO* dari metode *fuzzy regression* digunakan tool Minitab yang mana dengan menggunakan data input masing-masing pembagian data training seperti pada Tabel 4.16, Tabel 4.17, Tabel 4.18 dan Tabel 4.19 dan menghasilkan Persamaan (13), (14), (15) dan (16) dari hasil analisis pada tool minitab seperti pada Gambar 4.9, Gambar 4.10, Gambar 4.11 dan Gambar 4.12 berikut:



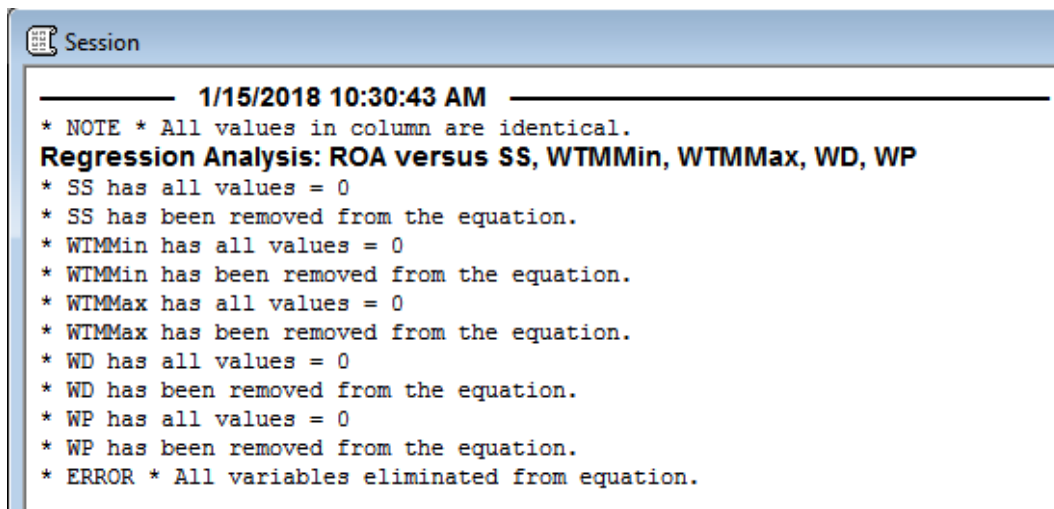
Gambar 4.9. Result ROAHIGH Regression Analysis with Minitab Tool



Gambar 4.10. Result ROAMED Regression Analysis with Minitab Tool



Gambar 4.11. Result ROALOW Regression Analysis with Minitab Tool



Gambar 4.12. Result ROANO Regression Analysis with Minitab Tool

$$ROAHIGH = 0.566 + 0.315 WD + 0.0692 WP + 0.0096 TMIN + 0.039 TMAX \quad (13)$$

$$ROAMED = 0.727 - 0.0014 WP + 0.151 TMIN - 0.00106 TMAX \quad (14)$$

$$ROALOW = 0.664 + 0.0737 WP - 0.050 TMIN + 0.035 TMAX \quad (15)$$

$$ROANO = 0 + 0 SS + 0 WD + 0 WP + 0 TMIN + 0 TMAX \quad (16)$$

Setelah mendapatkan model persamaan *fuzzy regression* dari data *training* kemudian sekarang men-*testing data* dengan data testing per masing-masing kriteria. Berikut adalah contoh proses mentesting data dari data *high anomaly fuzzy regression* yang diketahui dari *threshold high* seperti pada Tabel 4.20 berikut:

Tabel 4.20. Testing Data High Anomaly Fuzzy Regression

| Case | SS | WD | WP | TMIN | TMAX | ROA | Hasil Deteksi |
|------|----|------|------|------|------|------|---------------|
| 1 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |
| 2 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |
| 3 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |

| Case | SS | WD | WP | TMIN | TMAX | ROA | Hasil Deteksi |
|------|----|------|------|------|------|------|---------------|
| 4 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |
| 5 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |
| 6 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |
| 7 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |
| 8 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |
| 9 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |
| 10 | 1 | 0.74 | 0.66 | 0 | 0 | 0.84 | yes |
| 11 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 12 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 13 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 14 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 15 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 16 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 17 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 18 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 19 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 20 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 21 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 22 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 23 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 24 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 25 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 26 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 27 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 28 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 29 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 30 | 1 | 0.74 | 0.65 | 0 | 0 | 0.84 | yes |
| 31 | 1 | 0.1 | 0.1 | 1 | 1 | 0.65 | no |
| 32 | 1 | 0.1 | 0.1 | 1 | 1 | 0.65 | no |

Data terdeteksi benar *High Anomaly (Fuzzy Regression)* ketika antara data *predicted rate of anomaly* yang diperoleh dari model itu sama dengan data *ground truth* yang diperoleh dari *expert*. Dalam data testing *high anomaly (Fuzzy Regression)*, setelah diterapkan persamaan (13), *Case 1-30* memenuhi range *high*

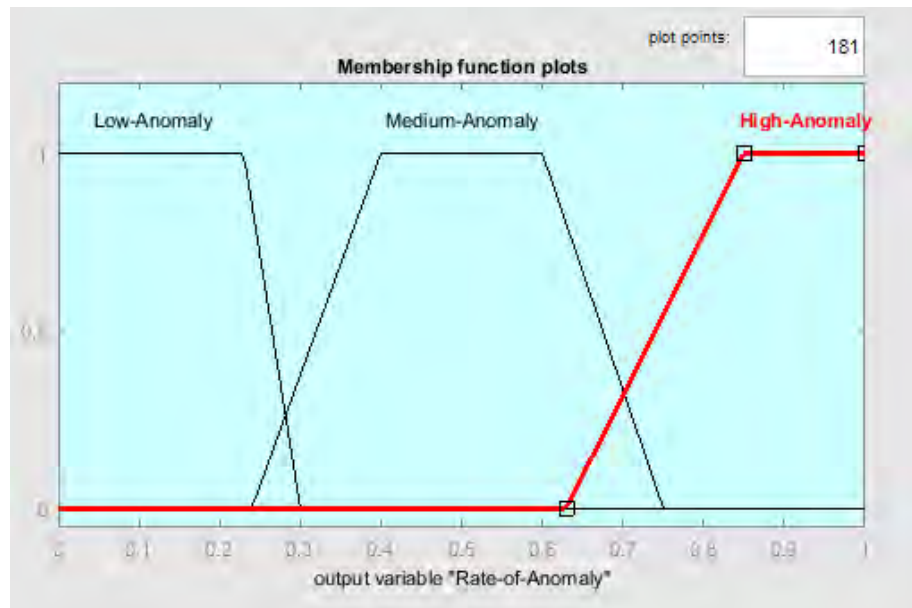
anomaly yang mana diantara range $0.8 \leq x \leq 1$, sedangkan *case* 31-32 tidak memenuhi range *high anomaly* (angka dibawah *threshold*). Maka dari itu terdapat 2 salah pendeteksian yang mana 5 dari medium. Dan alur pendeteksian ini diterapkan pada *Rate of Anomaly Medium*, *Rate of Anomaly Low* dan *Rate of Anomaly No (Fuzzy Regression)* juga. Dan hasilnya seperti pada Tabel 4.9.

Variabel atau atribut pada model *Multiple Linear Regression* dan *Fuzzy Regression* tidak sama dikarenakan pada saat pendeteksian anomali, terdapat atribut yang terkadang tidak ditemukan pada saat pendeteksian. Tidak semua atribut mengandung keempat indikator fuzzy (*High*, *Medium*, *Low* dan *No Anomaly*) dikarenakan bobot penilaian expert berbeda-beda seperti pada Tabel 4.21 berikut:

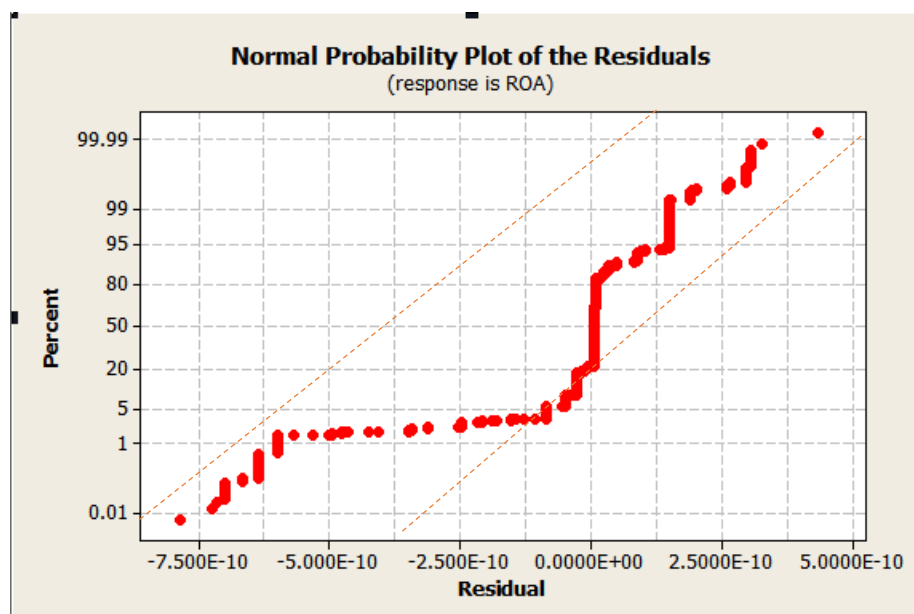
Tabel 4.21. Penggolongan Kategori Fuzzy

| Case | Skip Sequences | | | | Wrong Throughput Time MIN | | | | Wrong Throughput Time MAX | | | | Wrong Decision | | | | Wrong Patterns | | | | Rate of Anomaly | | | |
|-------|----------------|-----|-----|-----|---------------------------|-----|-----|-----|---------------------------|-----|-----|-----|----------------|-----|-----|-----|----------------|-----|-----|-----|-----------------|-----|-----|-----|
| | H | M | L | N | H | M | L | N | H | M | L | N | H | M | L | N | H | M | L | N | H | M | L | N |
| 1 | | | | √ | | √ | | | | √ | | | | | √ | | | √ | | | √ | | | |
| 2 | | √ | | | | | √ | | | | | √ | | | | √ | √ | | | | | √ | | |
| 3 | | | | √ | | √ | | | | | √ | | | | √ | | | | √ | √ | | | | |
| 4 | | | √ | | | √ | | | | | √ | | | | √ | | | √ | | | √ | | | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 65000 | | | | √ | | | | √ | | | √ | | | | | √ | | | √ | | | | √ | |

Terlihat pada Tabel 4.21 bahwa tanda merah merupakan contoh pada case training ketika pembuatan model *Fuzzy Regression* maupun *Multiple Linear Regression* bahwa terlihat tidak semua yang terdeteksi untuk pembangunan model mengandung kategori anomali *no anomaly*, *low anomaly*, *medium anomaly* dan *high anomaly*. Gambar 4.13 menunjukkan plot Fungsi Keanggotaan Fuzzy dari *Rate of Anomaly* yang terbentuk dari fungsi keanggotaan *Rate of Anomaly* dari data training yang berfungsi sebagai *threshold* penentuan kategori anomali, dalam hal ini yakni *low anomaly*, *medium anomaly* dan *high anomaly*. Untuk plot probabilitas normal pendeteksian anomali menggunakan *Fuzzy Regression* pada kategori *High Anomaly* dapat dilihat pada Gambar 4.14.



Gambar 4.13. Plot Membership Function Fuzzy Rate of Anomaly (ROA) High Anomaly



Gambar 4.14. Probability Plot of Detecting High Anomaly in Fuzzy High Anomaly

Tabel 4.22. Number of Cases of Training and Testing Fuzzy Regression

| <i>Class Label</i> | <i>Number of Cases for Training</i> | <i>Number of Cases for Testing</i> |
|--------------------|-------------------------------------|------------------------------------|
| High Anomaly | 48 | 32 |
| Medium Anomaly | 168 | 112 |
| Low Anomaly | 192 | 128 |
| No Anomaly | 38592 | 25728 |

Keanggotaan Fuzzy *Rate of Anomaly* dapat dilihat pada Persamaan (17), (18) dan (19):

RateofAnomaly[Low] =

$$\begin{array}{ll} 0 & ; \quad x < 0 \text{ atau } x > 0.3 \\ (x - 0)/(0.1 - 0) & ; \quad 0 < x < 0.1 \\ 1 & ; \quad 0.1 \leq x < 0.2 \\ (0.3 - x)/(0.3 - 0.2) & ; \quad 0.2 \leq x \leq 0.3 \end{array} \quad (17)$$

RateofAnomaly[Medium] =

$$\begin{array}{ll} 0 & ; \quad x < 0.23 \text{ atau } x > 0.75 \\ (x - 0.23)/(0.4 - 0.23) & ; \quad 0.23 \leq x < 0.4 \\ 1 & ; \quad 0.4 \leq x < 0.61 \\ (0.75 - x)/(0.75 - 0.61) & ; \quad 0.61 \leq x \leq 0.75 \end{array} \quad (18)$$

RateofAnomaly[High] =

$$\begin{array}{ll} 0 & ; \quad x < 0.61 \text{ atau } x > 1 \\ (x - 0.61)/(0.79 - 0.61) & ; \quad 0.61 \leq x < 0.79 \\ 1 & ; \quad 0.79 \leq x < 0.994 \\ (1 - x)/(1 - 0.994) & ; \quad x \geq 0.994 \end{array} \quad (19)$$

Nilai *Sensitivity* dan *Specificity* dari *Fuzzy Low Anomaly*, *Fuzzy Medium Anomaly*, dan *Fuzzy High Anomaly* dari pendeteksian memakai *Fuzzy Regression* dapat dilihat pada Tabel 4.17 dan Tabel 4.18.

Tabel 4.23. Confusion Matrix for Calculate Sensitivity Fuzzy Regression

| Expert System | High | Medium | Low | No Anomaly |
|------------------|----------------------|----------------------|----------------------|-------------------------|
| High | TH | FM | FL | FNa |
| Medium | FH | TM | FL | FNa |
| Low | FH | FM | TL | FNa |
| No Anomaly | FH | FM | FL | TNa |
| Sensitivity | $\frac{TH}{TH + FH}$ | $\frac{TM}{TM + FM}$ | $\frac{TL}{TL + FL}$ | $\frac{TAn}{TAn + FAn}$ |

Tabel 4.24. Confusion Matrix for Calculate Sensitivity Fuzzy Regression

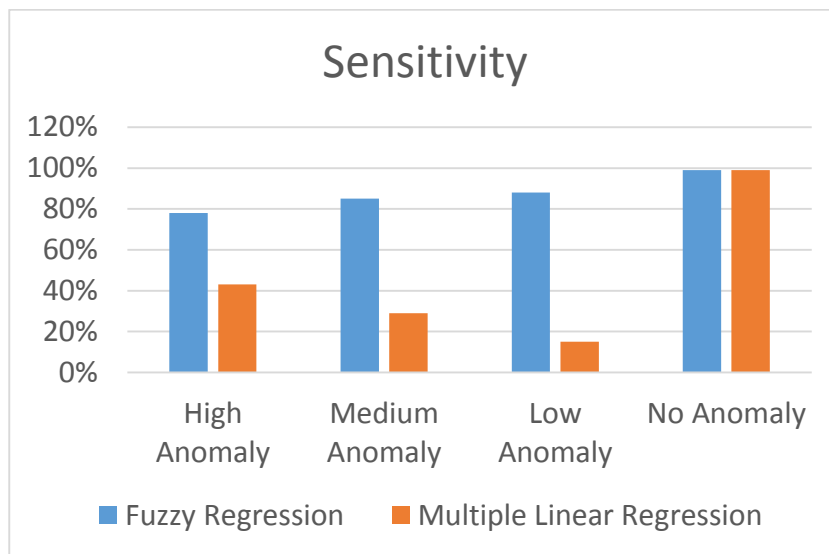
| Expert System | High | Medium | Low | No Anomaly |
|------------------|------------------------------------|------------------------------------|------------------------------------|----------------------------------|
| HIGH | TH | FH | FH | FH |
| MEDIUM | FM | TM | FM | FM |
| LOW | FL | FL | TL | FL |
| NO ANOMALY | FNa | FNa | FNa | TNa |
| SPECIFICITY | $\frac{TM + TL + TNa}{M + L + Na}$ | $\frac{TH + TL + TNa}{H + L + Na}$ | $\frac{TH + TM + TNa}{H + M + Na}$ | $\frac{TH + TM + TL}{H + M + L}$ |

Tabel 4.25. Result Detection Anomaly using Fuzzy Regression

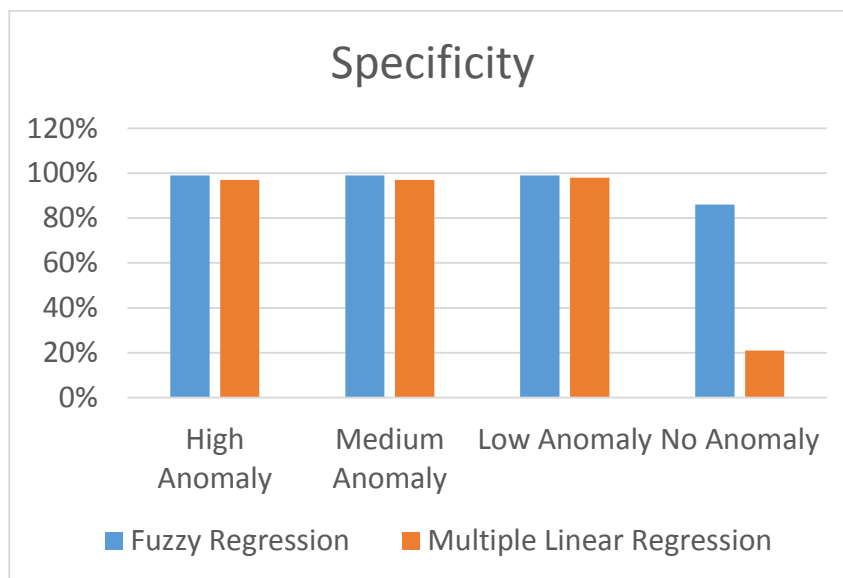
| EXPERT SYSTEM | HIGH | MEDIUM | LOW | NO ANOMALY |
|------------------|------|--------|-----|------------|
| HIGH | 30 | 2 | 0 | 0 |
| MEDIUM | 5 | 102 | 5 | 0 |
| LOW | 0 | 5 | 115 | 8 |
| NO ANOMALY | 3 | 10 | 10 | 25678 |
| SENSITIVITY | 78% | 85% | 88% | 99% |
| SPECIFICITY | 99% | 99% | 99% | 86% |

Untuk membandingkan hasil *Multiple Linear Regression* dan *Fuzzy Regression* dalam deteksi anomali atau fraud dilakukan evaluasi kinerja secara kuantitatif yakni dengan membandingkan nilai *sensitivity* dan *specitivity*. Dalam kasus pendeteksian anomali pada PT.Terminal Peti Kemas Surabaya ini, tidak ada secara khusus penanganan kasus imbalanced. Data akan langsung dideteksi anomali meski adanya ketidakseimbangan antara jumlah yang terdeteksi anomali dan yang tidak terdeteksi anomali (1:10 Contohnya).

Grafik perbandingan nilai *Sensitivity* dan *Specificity* antara Fuzzy Regression dan Multiple Linear Regression dari pendeteksian anomali pada PT. Terminal Peti Kemas (TPS) dapat dilihat pada Gambar 4.15 dan Gambar 4.16.



Gambar 4.15. *Sensitivity Value of FR and MLR*



Gambar 4.16. *Specificity Value of FR and MLR*

BAB V

KESIMPULAN

Kesimpulan dari penelitian ini adalah *Fuzzy Regression* adalah teknik baru untuk mengetahui tingkat prediksi kecurangan secara lebih akurat daripada regresi linier berganda, langkah pertama mengklasifikasikan kasus mana yang memiliki anomali rendah, anomali sedang dan anomali tinggi kemudian menemukan prediksi tingkat anomali dari Semua nilai anomali atribut yang nilainya diskrit (Rendah, Sedang dan Tinggi). Hasil penelitian menunjukkan bahwa regresi fuzzy lebih tepat untuk mendeteksi anomali daripada regresi linier berganda. Terbukti dengan sensitivitas dan spesifisitas *Fuzzy Regression* untuk anomali yang tinggi adalah 78% dan 99%, sensitivitas dan spesifisitas *Fuzzy Regression* untuk anomali medium adalah 85% dan 99%, sensitivitas dan spesifisitas *Fuzzy Regression* untuk anomali rendah adalah 88% dan 99% dan sensitivitas dan spesifisitas *Fuzzy Regression* untuk tidak anomali adalah 99% dan 86%.

[Halaman ini sengaja dikosongkan]

DAFTAR PUSTAKA

- [1] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen and X. Sun, "The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature", *Decision Support Systems*, vol. 50, no. 3, (2010), pp. 559-569.
- [2] I. Amara, A. B. Amar and A. Jarboui, "Detection of Fraud in Financial Statements: French Companies as a Case Study", *International Journal of Academic Research in Accounting, Finance, and Management Sciences*, vol. 3, no. 3, (2013), pp. 44-55.
- [3] S. Huda, R. Sarno, T. Ahmad, H. A. Santosa, "Identification of Process-based Fraud Patterns in Credit Application", *International Conference on Information and Communication Technology (ICoICT)*, (2014) June 28-29.
- [4] R. Sarno, D.R. Dewandono, T. Ahmad, M. F. Naufal and F. Sinaga, "Hybrid Association Rule Learning and Process Mining for Fraud Detection", *IAENG International Journal of Computer Science*, vol. 42, no. 2, (2015), pp. 59-72.
- [5] S. Huda, R. Sarno, and T. Ahmad, "Fuzzy MADM approach for Rating of Process-based Fraud", *Journal ICT. Research Application*, vol. 9, no. 2, (2015), pp. 111-128.
- [6] R. Sarno, A.B. Sanjoyo, I. Mukhlash and M.H. Astuti, "Petri Net Model of ERP Business Process Variations for Small and Medium Enterprises", *Journal of Theoretical and Applied Information Technology*, vol. 54, no. 1, (2013), pp. 31-38.
- [7] R. Sarno, and C. A. Djeni, "Developing a Workflow Management System For Enterprise Resource Planning", *Journal of Theoretical and Applied Information Technology*, vol. 72, no. 3, (2015), pp. 412-421.
- [8] R. Sarno, P. L. I. Sari, H. Ginardi, D. Sunaryono, I. Mukhlash, "Decision Mining For Multi Choice Workflow Patterns", *International Conference on Computer, Control, and Its Application*, (2013) November 19-21.
- [9] M. Jans, M. Alles, and M. Vasarhelyi, "The Case for Process Mining in Auditing: Sources of Value Added and Areas of Application", *International Journal of Accounting Information Systems*, vol. 14 no. 1, (2013), pp. 1-20.
- [10] W. M. P. van der Aalst, "Discovery, Conformance, and Enhancement of Business Processes", *Springer*, pp. 7-8, (2010).
- [11] Stoop, J.J., *Process Mining, and Fraud Detection*, Thesis, Business Information Technology Department, Twente University, Enschede, Netherlands, 2012.
- [12] Dewandono, D.R., *Process Sequence Mining For Fraud Detection Using CEP*, Thesis, Informatics Department, Institut Teknologi Sepuluh Nopember, Surabaya, 2013.
- [13] Bardossy, A., I. Bogardi and L. Duckstein. (1990) "Fuzzy regression in hydrology", *Water Resources Research* 26, 1497-1508.
- [14] Celmiņš, A. (1987) "Least squares model fitting to fuzzy vector data," *Fuzzy Sets and Systems*, 22(3), 245-269

- [15] Chang, Y.-H. O. and B. M. Ayyub. (2001) "Fuzzy regression methods – a comparative assessment," *Fuzzy Sets and Systems*, 119(2), 187-203
- [16] Hojati, M., C. R. Bector and K. Smimou. (2004) "A simple method for computation of fuzzy linear regression," *European Journal of Operational Research* (forthcoming)
- [17] Hong, D. H., J-K. Song and H.Y. Do. (2001) "Fuzzy least-squares linear regression analysis using shape preserving operations," *Information Sciences* 138 185-193
- [18] Ishibuchi, H. (1992) "Fuzzy regression analysis," *Fuzzy Theory and Systems*, 4, 137-148
- [19] Kao, C. and C-L Chyu. (2003) "Least-squares estimates in fuzzy regression analysis," *European Journal of Operational Research* 148, 426-435
- [20] Kim, K. J., H. Moskowitz and M. Koksalan. (1996) "Fuzzy versus statistical linear regression," *European Journal of Operational Research*, 92(2) 417-434
- [21] Koissi, M-C, and A. F. Shapiro. (2005) "Fuzzy formulation of Lee-Carter mortality model," working paper.
- [22] McCauley-Bell, P. and H. Wang. (1997) "Fuzzy linear regression models for assessing risks of cumulative trauma disorders," *Fuzzy Sets and Systems*, 92(3), 317-340
- [23] Peters, G. (1994) "Fuzzy linear regression with fuzzy intervals," *Fuzzy Sets and Systems*, 63(1), 45-55
- [24] Sánchez, J. de A., and A. T. Gómez. (2003a) "Applications Of Fuzzy Regression In Actuarial Analysis," *JRI* 2003, 70(4), 665-699
- [25] Sánchez, J. de A., and A. T. Gómez. (2003b) "Estimating a term structure of interest rates for fuzzy financial pricing by using fuzzy regression methods," *Fuzzy Sets and Systems*, 139(2), 313-331
- [26] Sánchez, J. de A., and A. T. Gómez. (2004) "Estimating a fuzzy term structure of interest rates using fuzzy regression techniques," *European Journal of Operational Research*, 804–818
- [27] Savic, D. A., and W. Pedrycz. (1991) "Evaluation of fuzzy linear regression models," *Fuzzy Sets and Systems*, 39(1), 51-63
- [28] Tanaka, H., Uejima, S., and Asai, K. (1982) "Linear regression analysis with the fuzzy model," *IEEE Transactions on Systems, Man, and Cybernetics*, 12(6), 903-907.
- [29] Wang, H.-F., and R.-C. Tsaur. (2000) "Insight of a fuzzy regression model," *Fuzzy Sets and Systems*, 112(3), 355-369
- [30] Wünsche, A. and W. Näther. (2002) "Least-squares fuzzy regression with fuzzy random variables," *Fuzzy Sets and Systems*, 130(1), 43-50.
- [31] T. R. Berry-Stölzle, M.-C. Koissi, and A. F. Shapiro. (2010) "Detecting fuzzy relationships in regression models: The case of insurer solvency surveillance in Germany," *Insurance: Mathematics and Economics*, vol. 46, no. 3, pp. 554–567,
- [32] A. F. Shapiro. (2005). "Fuzzy Regression Models", Penn State University,

Smeal College of Business, University Park.

- [33] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento, “A (sub)graph isomorphism algorithm for matching large graphs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 10, pp. 1367–1372, Oct. 2004.
- [34] C. Nabti and H. Seba, “Subgraph Isomorphism Search in Massive Graph Databases,” *Proceedings of the International Conference on Internet of Things and Big Data*, 2016.
- [35] B. D. McKay, “Graph Isomorphism,” *Encyclopedia of Algorithms*, pp. 373–376, 2008.
- [36] W. M. P. van der Aalst, “Workflow Patterns,” *Encyclopedia of Database Systems*, pp. 1–2, 2016.

[Halaman ini sengaja dikosongkan]

DAFTAR SINGKATAN

| | |
|--------------|---|
| BPM | : <i>Business Process Management</i> atau Manajemen Proses Bisnis |
| BPMN | : <i>Business Process Management and Notation</i> |
| CFP | : <i>Control Flow Patterns</i> |
| CPN | : <i>Coloured Petri Net</i> |
| CSV | : <i>Comma Separated Values</i> |
| FLSR | : <i>Fuzzy Least-Square Regression</i> |
| FR | : <i>Fuzzy Regression</i> |
| LS | : <i>Least Square</i> |
| MLR | : <i>Multiple Linier Regression</i> |
| MFs | : <i>Membership Fuzzy</i> |
| MXML | : <i>Mining eXtensible Markup Language</i> |
| OLS | : <i>Ordinary Least Squares</i> |
| ROA | : <i>Rate of Anomaly</i> |
| SF | : <i>Straight Forward</i> |
| SOP | : <i>Standard Operating Procedure</i> |
| STFNs | : <i>Symmetrical Triangular Fuzzy Numbers</i> |
| TFNs | : <i>Triangular Fuzzy Numbers</i> |
| TSIR | : <i>Term Structure of interest rates</i> |
| XES | : <i>eXtensible Event Stream</i> |
| YAWL | : <i>Yet Another Workflow Language</i> |

[Halaman ini sengaja dikosongkan]

DAFTAR ISTILAH

Anomaly

Suatu penyimpangan yang terjadi pada suatu tingkah laku proses bisnis yang dapat menimbulkan kecurigaan bahwa proses tersebut terindikasi terjadi *fraud*

Attribute Name

Nama kriteria yang menggambarkan jenis pelanggaran yang terjadi di dalam proses.

Attribute Value

Nilai penyimpangan atau anomali dari SOP

Activity

Merupakan bagian dari case yang merupakan sub proses dalam pembuatan suatu barang atau dalam suatu proses tertentu.

Activity Lifespan

Selisih waktu akhir dan mulai sebuah aktivitas.

Alpha, Alpha+, Alpha++

Salah satu algoritma yang digunakan untuk melakukan proses *discovery*.

AND

Parallel AND terjadi jika parallel split *pattern* muncul. Parallel split *pattern* didefinisikan sebagai mekanisme yang memungkinkan dua kegiatan yang berbeda dilakukan secara bersamaan. Sifat dasar dari pola ini sendiri adalah semua aktivitas yang ada di percabangan harus dijalankan, baik itu dijalankan secara bersamaan atau secara bergantian.

Business Process Model

Rangkaian dari aktivitas dengan mengikuti standar proses tertentu dengan penggambaran model bisa berdasarkan dari *event log* maupun penggambaran secara manual

Case

Suatu kasus tertentu yang ada pada *event log*. Kasus tertentu tersebut dapat berupa suatu kasus dalam memproduksi suatu barang tertentu, karena *event log* dapat terdiri dari catatan dari proses eksekusi pembuatan banyak barang atau proses eksekusi dari banyak kasus proses.

Completeness

Suatu kondisi dimana *event log* menyimpan seluruh perilaku yang bisa dieksekusi pada proses bisnis.

CPN (Coloured Petri Net)

Kakas bantu yang digunakan untuk mengedit, mensimulasi dan menganalisis coloured petri net.

Discovery

Salah satu teknik *process mining* yang bertujuan untuk mendapatkan proses model dengan menggali informasi dari *event log*.

Event log

Suatu set proses eksekusi yang mengambil dari data aktivitas proses bisnis yang dilakukan dalam konteks tertentu.

Fuzzy miner

Salah satu algoritma yang digunakan untuk melakukan proses *discovery*.

OR

Conditional OR digunakan ketika *multiple choice pattern* muncul. *Multiple choice pattern* pemilihan satu atau lebih aktivitas dalam percabangan untuk dijalankan. Dalam *multiple choice pattern* satu aktivitas dapat dijalankan sendiri tanpa harus menjalankan aktivitas lain yang ada dipercabangan, atau juga dapat menjalankan beberapa aktivitas baik secara bersamaan maupun tidak.

Paralel

Suatu rangkaian proses bisnis dimana eksekusi dari aktivitas yang terdapat dalam rangkaian tersebut dapat dilakukan secara bersamaan maupun berurutan.

Process discovery

Salah satu proses yang paling menantang dari rangkaian *process mining*, tujuan dari proses ini adalah untuk membentuk model dengan cara menggali informasi dari data yang tercatat dalam suatu *event log*.

Process mining

Teknik yang dapat digunakan untuk mendapatkan model sesungguhnya dari proses bisnis yang terjadi dalam sebuah sistem informasi berdasarkan data yang berasal dari *event log*.

ProM

Kakas bantu yang digunakan untuk menganalisa dan menggambarkan proses berdasarkan data. Terdapat dua versi dari PROM yaitu PROM 5.2 dan PROM 6, dimana PROM 5.2 yang digunakan pada modul ini. Terdapat beberapa pilihan algoritma pada PROM yang dapat digunakan untuk menganalisa proses

Proses bisnis

Rangkaian dari aktivitas yang dibuat untuk menghasilkan keluaran spesifik dengan tujuan tertentu.

Rate of Anomaly

Nilai anomali keseluruhan dari setiap kasus diambil dari nilai bobot per atribut

Sensitivity

Perhitungan proporsi bagian positif yang dapat diidentifikasi dengan benar

Skip Activity

Jenis anomali yang terjadi ketika salah satu aktivitas di SOP sengaja dilewati atau tidak dikerjakan

Skip Sequence

Skip yang terjadi pada aktivitas sequence

SOP (Standard Operating Procedure)

Suatu set instruksi (perintah kerja) terperinci dan tertulis yang harus diikuti demi mencapai keseragaman dalam menjalankan suatu pekerjaan tertentu.

Specificity

Perhitungan proporsi bagian negatif yang dapat diidentifikasi dengan benar

Timestamp

Suatu informasi yang menyimpan data tanggal dan waktu suatu kejadian dilakukan.

Trace

Alur dari aktivitas yang dijalankan dalam suatu proses.

Waiting Time

Waiting time adalah waktu yang dibutuhkan suatu proses selama menunggu di *ready queue*

Weight of Attribute

Pembobotan Jenis Pelanggaran / Anomali pada setiap *attribute value*

Wrong Throughput Time

Jenis anomali yang terjadi ketika waktu pengerjaan aktivitas melebihi rata-rata waktu pengerjaan di aktivitas tersebut dan batas toleransi yang telah ditentukan

Wrong Throughput Time Minimum

Wrong Throughput Time yang waktu pengerjaan aktivitas lebih kecil dari batas toleransi bawah yang ditentukan

Wrong Throughput Time Maximum

Wrong Throughput Time yang waktu pengerjaan aktivitasnya lebih besar dari batas toleransi atas yang ditentukan

Wrong Pattern

Jenis anomali yang terjadi ketika urutan aktivitas yang dikerjakan tidak sesuai dengan SOP

Wrong Decision

Jenis anomali yang terjadi ketika alur keputusan yang diambil salah atau tidak sesuai dengan ketentuan SOP

XOR

Single Choice XOR terjadi jika titik dalam proses alur kerja di mana satu cabang dibagi menjadi dua atau lebih tetapi trace hanya dapat memilih salah satu cabang saja

YAWL (*Yet Another Workflow Language*)

Salah satu bahasa pemodelan proses bisnis.

[Halaman ini sengaja dikosongkan]

DAFTAR INDEKS

| | |
|------------------------------------|---|
| AND, 77 | Petri Net, 75, 78 |
| <i>case</i> , 77 | <i>process mining</i> , 78, 85 |
| <i>event log</i> , 78 | proses bisnis, 78 |
| Manajemen Proses Bisnis, 75 | <i>Standard Operating Procedure</i> , 78 |
| OR, 78 | XOR, 78 |

[Halaman ini sengaja dikosongkan]

BIOGRAFI PENULIS



Dewi Rahmawati, S.Kom.

Lahir di Malang, Jawa Timur pada tanggal 9 Maret 1994. Saat ini sedang menempuh pendidikan program magister di Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember serta menjadi dosen di PIKTI ITS, Project Evaluations and Documentations di PT Profio Teknova Indonesia dan Trainer Nasional Indonesia Next bidang “*Microsoft Office Specialist*” di Medan dan Surabaya. Bidang penelitian dari penyusun adalah *process mining* dan deteksi *fraud*. Penyusun dapat dihubungi melalui e-mail: **dewirahmaw4ti@gmail.com.**